Arrays, Methods and Kits for Preparation and Use of Syntenic Genomic Arrays for Diagnostics and Toxicology

5

Technical Field

This invention relates to molecular biology, genetic diagnostics and array, or "chip" or "biochip" technology. In particular, the invention provides methods for determining chromosomal abnormalities in a cell, an organism, or a cell population, such as in cancer cells

10   or in embryonic cells, and for comparing chromosome abnormalities of a plurality of different species of organisms with respect to defects that affect a chromosomal syntenic strand that are homologous in the plurality of organisms. The invention provides methods, arrays produced by the methods, and kits for analysis of nucleic acids, for diagnosis, prognosis, and toxicology.

15   **Background**

Genomic DNA array based chips have the potential to solve many of the limitations of traditional whole chromosome analysis methods, which rely on hybridization of samples to individual metaphase chromosomes. In contrast to metaphase hybridization, in which the immobilized genomic DNA is a metaphase spread, array-based hybridization uses

20   immobilized nucleic acids arranged as an array on a biochip or an array platform. The array hybridization approach can provide DNA sequence copy number information across the entire genome in a single, timely, cost-effective and sensitive procedure, the resolution of which is primarily dependent upon the number, size and map positions of the DNA elements within the array. Typically, bacterial artificial chromosomes, or BACs, which can each

25   accommodate on average about 150 kilobases (kb) of cloned genomic DNA, are used in the production of the array.

While array genome profiling represents a revolutionary progression in genetic testing, certain aspects of the technique continue to limit performance. In many cases, application and immobilization of a nucleic acid probe to a substrate produce uneven deposit

30   of the genomic nucleic acid across the surface of the spot, yielding samples that are not uniform when viewed under magnification, such as in a microscope. Further, rare but

1

troublesome incomplete removal of non-specifically bound nucleic acid test and reference sample to areas of the substrate can lead to complexity of analysis.

Summary of Embodiments of the Invention

A problem in toxicology and in environmental analyses is comparing results obtained with an experimental animal to those that might be observed in a human, or affects a human, or might affect a human.

A featured embodiment of the present invention is a nucleic acid array that has a plurality of immobilized elements in an array, the elements at addressable locations on a substrate, the elements being "spots" or patches of nucleic acid deposited on the substrate. With the array of the featured invention, the plurality of elements comprise nucleic acid sequences from a chromosome syntenic strand from each of a plurality of organisms, and a first set of the elements and a second set of the nucleic acid sequences in the elements are chosen so that the first set and the second set are from different species of organism, and the elements that are from a syntenic chromosome of a first species of organism have nucleic acid sequences that are homologous to the nucleic acid sequences that are from the syntenic chromosome of a second species of organism.

The term array as used herein implies a plurality which is in this case a very large number of elements on a surface, for example, at least 10, or at least 100, or at least 200, or at least 1,000. The elements are generally non-identical, however duplicate or triplicate spots are used for statistical significance. Each non-identical element contains a nucleic acid having a nucleic acid sequence that is a marker for that element, and that distinguishes it from other non-identical elements on the same surface. The term array further implies an orderly arrangement such that each spot is "addressible", i.e., has a known location and a known nucleic acid sequence content.

Calibration spots are included in embodiments of each array herein, as described in U.S. patent application serial number 10/112,657 filed March 27, 2002. In one embodiment, a calibration spot contains a sample of all or substantially all of the non-identical sequences in the other elements of the array. In another embodiment, various concentrations or dilutions of a calibration spot are included in each array. In yet another embodiment, a calibration spot includes a known quantity of a calibration molecule, which can be pre-labeled or it not pre-labeled, e.g., the calibration molecule can be obtained from a species other than that of interest in the remainder of the elements of the array, such as Escherichia coli DNA or Xenopus laevis DNA used as a calibration spot for syntenic arrays having nucleic acid elements of human and mouse DNA; in other embodiments, a calibration

2

molecule can be a synthetic nucleic acid having a naturally occurring or a non-natural nucleic acid sequence.

In general, the elements of the syntenic array of the embodiments herein has nucleic acid that is cloned genomic DNA. For example, the cloned genomic DNA is carried on a
5   vector selected from the group of vectors consisting of yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), mammalian artificial chromosomes (MACs), and phage P1 artificial chromosomes (PACs).

Further, in general at least one of the species of organism is a mammal. However may other embodiments of "syntenic arrays" are envisioned as within the scope of the
10   invention, such as those comparing any two different organisms, such as species of crop plants, or freshwater fish. In various embodiments, at least one of the organisms is a mamal, for example, at least of the organisms is a human. Further, at least one organism is selected from the group consisting of rodents, non-human primates, marine mammals, freshwater mammals, lagomorphs, porcines, bovines, carnivore, caprines, equines, amphibia, fish, and
15   insects. More specifically, at least one organism is selected from the group consisting of a gorilla, a chimpanzee, a monkey, a dog, a hamster, a mouse, a rat, a rabbit, a guinea pig, a sheep, a goat, a swine, a cow, a horse, a frog, a toad, a zebra fish, and a fly. In an exemplary array, the species of organism are human and mouse. With a human-mouse array, effects on genotoxicity of, for examples, experimental mice or mouse cells can be compared with
20   effects on genotoxicity on human cells, for example, in culture. In yet another exemplary array, the species of organism are human and a wild animal.

Any of the non-human organisms are, in related embodiments, transgenic, i.e., a transgenic mouse that is used for screening compositions to obtain a particular novel activity capable of remediating a particular phenotype is then used as a source of DNA, for
25   hybridization to the syntenic arrays ("chips") herein. In this manner, any screen of organisms or cells can be further analyzed for genotoxicity, in particular, to identify compounds that do not result in chromosomal abnormalities. In addition, the animals herein can have a "model" disease, which as used herein means a disease in an animal that is induced or is present genetically, and that has symptoms and a phenotype similar to that of a disease of humans or
30   a non-human animal, and that is useful for analysis of agents capable of remediating that disease.

At least one element of nucleic acid of the first species is at least about 50% homologous to at least one element of nucleic acid of the second species. Further, at least one element of nucleic acid of the first species is at least about 70% homologous to at least

3

one element of nucleic acid of the second species, or about 80%, about 85%, about 90%, about 95% or about 99% homologous to at least one element of nucleic acid of the second species.

5   The array elements include nucleic acid sequences that are representative of at least one chromosome of at least one species. In a related embodiment, the array elements include nucleic acid sequences that are representative of a genome of at least one of the species. Representative of at least one chromosome means that at least three, four, five or more elements are present in the array that contain sequences from different points along the chromosome, such that data obtained from a hybridization of a nucleic acid sample to the

10   array can be plotted from the p-terminus to the q-terminus for that chromosome. Similarly, representative of at least one genome means that all of the chromosomes within that genome are represented, e.g., in a human array, elements are present in the array that contain sequences from different points along each of all of the 22 autosomes and the X and the Y chromosomes. In a related embodiment, the elements of the array include nucleic acid

15   sequences representative of genomes of at least two species.

For all of the above syntenic arrays or the methods below, further embodiments include providing the arrays as multi-array surfaces. The multi-array surfaces have a plurality of any of the above arrays on a single substrate. Each of the plurality of arrays is printed on the surface in a pattern that is non-contiguous with others, so that a plurality of

20   hybridizations can be carried out on the same substrate. For example, two arrays can be printed with each one at each end of a glass slide, or three can be printed in a linear arrangment with one array at each end and one in the middle. The individual arrays within the plurality can further be separated by hydrophobic strips such as a Teflon strip; alternatively or in addition, or barriers ("dykes") or raised portions of a surface such as a slide

25   can be custom designed to be present prior to printing the array, or added later. In additional embodiments of the arrays herein, a viscosity-enhancing solute such as a dextran or a polyethyelene glycol can be added to the hybridization buffer, to enhance separation of the plurality of hybridizations being performed. Finally, a cover such as a cover slip can be separately applied to each hybridization mixture on each array within the multi-array surface.

30   Another featured embodiment of the invention is a method of measuring genotoxicity of a composition to a cell of a species of organism, the method comprising contacting a test cell or a cell population of a first species with the composition; obtaining a sample of nucleic acid from the contacted test cell or population; and analyzing the genome of the sample nucleic acid for abnormalities by hybridizing the nucleic acid to an array of syntenic nucleic

acid immobilized at addressible locations on a substrate, the syntenic array having elements of sequences of nucleic acid from the genome of the first species, and having elements of sequences of syntenic nucleic acid from the genome of at least a second species of organism.

In general, the second species is a human. Further, contacting the test cell with the compositions is, in some embodiments, adding the composition to a cell or a population of the first species in culture. Use of cell culture makes it possible, for example, to determine the effects of a variety of agents on human cells.

In an alternative embodiment, contacting the test cell or population is treating the cell of the first species in vivo, i.e., treating the intact organism, generally a multicellular organism. Thus. treating the cell in vivo is administering the composition by a route selected from the group of administering orally, topically, transdermally, and injecting. Injecting can be intravenous, subcutaneous, intraperitoneal, and any other standard route. Treating can also be, depending on the species, by rectal, intravaginal, intrathecal administration.

In an embodiment of the method, the first species is a subject exposed to the composition in a natural environment. The subject can be a "wild" organism such as a wild animal or plant, or the subject can be an experimental animal as is used in a laboratory that has been placed in the wild in order to measure for a presence of a genotoxic agent. In a different embodiment, a subject can be a human or animal patient that has been inadvertantly or intentionally exposed to an agent, and the agent was not previously known to be genotoxic.

Accordingly in the method, the first species which is the test species is selected from the group consisting of gorilla, chimpanzee, monkey, dog, hamster, mouse, rat, rabbit, guinea pig, sheep, goat, swine, cow, horse, frog, toad, fish, and insect. An example of the first species is a non-human transgenic experimental animal. Another example of the first species is an animal having a model disease, for example, a mouse having experimental allergic encephalomyelitis (EAE), or a non-obese diabetic mouse (NOD), or a mouse treated with streptozotocin to induce diabetes. However as the method can be practiced following a variety of different circumstances, the first species can be a human cell or a human subject.

Analyzing the genome of the contacted organism according to the method further comprises comparing hybridization of nucleic acid from the test cell to hybridization of nucleic acid from a reference cell or cell population. The reference cell is from the first species; alternatively, the reference cell is from the second species, or from a third species. In some embodiments, the reference cell or cell population is not administered the composition, and is otherwise identical to the test cell. In general in any of the embodiments

of the methods herein, the nucleic acid of the elements immobilized in the array on the substrate is cloned DNA.

The method generally includes, prior to hybridization, labeling separately each of the test cell nucleic acid and the reference cell with a first detectable label and a second

5   detectable label. For example, the first and second labels are fluorescent dyes, and the dyes have different emission spectra.

The method further includes, after labeling, preparing a first mixture of the test cell nucleic acid labeled with the first label or dye and the reference cell nucleic acid labeled with the second label or dye, and preparing a second mixture of the test cell nucleic acid labeled

10   with the second label or dye and the reference cell nucleic acid labeled with the first label or dye, and separately hybridizing each of the first mixture and the second mixture to iterations of the syntenic array.

The method further includes comparing the genome of the test cell by normalizing a ratio of extent of hybridization of the first and second labels or dyes to each element for each

15   of the first and second mixtures. The method further includes plotting the resulting set of ratios as a function of the location of each of the nucleic acids as a distance along a chromosome from the p-terminus to the q-terminus. Comparing the genome is further identifying a chromosome of the test organism having a chromosomal abnormality. A chromosomal abnormality includes an increase or decrease in copy number, such as a

20   deletion or an amplification, and also includes a translocation, an inversion, and an insertion, including a presence of a nucleic acid sequence not previously characterized at a location along the chromosome.

The method further includes identifying a chromosomal location along the chromosome of the abnormality in the test sample, with respect to the array of immobilized

25   elements of the test organism. The method further involves, in cases in which the first species is non-human and the syntenic array comprises elements of the human genome, determining an homologous chromosome and chromosomal location of the abnormality in the human genome. In this context, the term, "homologous chromosome" means a chromosome of one species having substantial nucleic acid homology with a chromosome of

30   another species. Another term often used to describe sequence homology in a different organism is "orthologous".

The method further can be used, by comparing an amount of chromosomal abnormalities in the test sample nucleic acid to chromosomal abnormalities in the reference sample nucleic acid, as an indication of an extent of genotoxicity of the chemical

composition. Further, comparisons of chromosomal locations of the abnormalities in the test species and in another species of organism, for example, in a human, can be made.

The method is used to analyze genotoxicity of the composition which is exemplified by and not limited to: a hazardous occupational compound, a chemical weapon, airborne dust,
5   photochemical smog, a natural product, a cosmetic, a food additive, an agricultural product, an industrial compound, a new chemical entity, a lead compound, a pharmaceutical product, sewage, and an environmental sample, and an extract or preparation of any of these agents or components of any of these agents.

Another featured embodiment of the invention is a method of identifying a presence of
10   a genotoxic agent for a cell of a species of organism in a natural environment, the method comprising obtain a test cell or a cell population of a first species of organism in the environment; obtaining a sample of nucleic acid from the test cell or population; and analyzing the genome of the sample nucleic acid for abnormalities by hybridizing the nucleic acid to an array of syntenic nucleic acid immobilized at addressible locations on a substrate,
15   the syntenic array having elements of sequences of nucleic acid from the genome of the first species, and having elements of sequences of syntenic nucleic acid from the genome of at least a second species of organism. The first species may be a feral organism, i.e., one having a life cycle in the natural environment, or may be a laboratory strain that has been placed in the environment. Alternatively, the first species can be a human exposed to an agent
20   inadvertently, such as a human subjected to an ocupational hazard which may be, for example, a physical force or a chemical composition in the occupational environment.

Also among the embodiments of the invention provided herein is a kit for use of the method according to any of the above, comprising a syntenic array having immobilized nucleic acid elements with nucleotide sequences from genomes of a plurality of species of
25   organism, and a container. The kit can further include any of the reagents, e.g., a plurality of detectible labels, and/or a polymerase for amplification of nucleic acids, or a computer program for obtaining and/or analyzing data obtained from hybridization to the array, and instructions for use.

Yet another featured embodiment provided herein is a method of identifying the
30   presence and location of chromosomal abnormalities in cells of a subject during progression of a disease, the method comprising obtaining a nucleic acid sample from the cells affected by the disease; and analyzing the sample for chromosomal abnormalities by hybridizing the sample to elements of a first syntenic nucleic acid array having nucleic acid from the genome of a first species, and further hybridizing the sample to elements of a second syntenic nucleic

acid array having nucleic acid from the genome of a second species, the elements of the first and second arrays being immobilized on a substrate. Accordingly, the method involves obtaining an additional nucleic acid sample of cells from the subject at a time point representing a different stage of progression of the disease. Progression of the disease can be

5   determined by comparing chromosomal abnormalities in samples from a plurality of different time points, i.e., at least two different time points.

For studying progression of a disease by the method herein, a first species or a second species is human; in this embodiment or in an alternative embodiment, the disease is an animal model of a human disease, i.e., the animal model is a strain of experimental animal, or

10   is an experimental animal treated to produce a disease condition that is useful for study of a human disease. Diseases such as lung cancer, mesothelioma, adenocarcinoma, and prostate cancer have "animal models" well known to one of ordinary skill in the art of experimental approaches to study of human disease.

Typically, the disease is a cancer, for example, the disease is a solid tumor, a blood

15   proliferative condition. The disease is selected from the group of cancers of skin, lung, breast, head and neck, prostate, ovary, brain, leukemia, gastric, stomach, esophagous, pancreas, and lymphoma. The disease is a stage I cancer; alternatively, the cancer is selected from stage II, III and IV cancers. Accordingly in certain embodiments, the cancer is metastatic.

20   Also provided herein is a method of preparing an array of a plurality of elements of a class of biological macromolecules immobilized on a substrate, each element of the array having a uniform distribution of the macromolecules, the method comprising contacting the substrate with the macromolecules in a composition comprising a buffer of effective ionic strength such that a uniform distribution of the macromolecules is obtained across the surface

25   of the element. The class of biological macromolecules is selected from the group consisting of nucleic acids, proteins, lipids, and carbohydrates. For example, the nucleic acids are DNA, for example, the nucleic acids are genomic DNA clones. In a related embodiment, the clones comprise an artificial chromosome library of a genome. The artificial chromosomes are selected from yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs),

30   mammalian artificial chromosomes (MACs), and phage P1 artificial chromosomes (PACs). In a preferred embodiment, the artificial chromosomes are BACs.

Accordingly, the ionic strength of the buffer for the contacting step is at least 100 mM, for example, the ionic strength is at least 150 mM. Further, the ionic strength is less than 1.0 M, for example, the ionic strength is less than 500 mM. The buffer in certain

embodiments comprises an organic ion, for example, the buffer is TRIS *N*-
[tris(hydroxymethyl)methyl]glycine (TRIS); 4-(2-hydroxyethyl)-1-piperazineethanesulfonic
acid (HEPES); 3-(N-morpholino) propanesulfonic acid (MOPS); or 4-
morpholineethanesulfonic acid (MES). In a preferred embodiment, the buffer is TRIS.

5   Alternatively, the buffer comprises an inorganic ion, for example, the buffer comprises a
phosphate ion. In general, the buffer further comprises EDTA. The buffer has a pH of at
least about 7, and a pH that is less than about 9. The method further includes drying the array
of elements on the substrate. For any of the methods or arrays herein, the substrate is
selected from the group consisting of: glass, paper, ceramics, quartz, metals, plastics, nylon,

10   teflon, silicones, and cellulose acetate. Typically, the substrate is a glass slide.

       Embodiments of any of the methods herein can further include providing the arrays as
a multi-array surface, and conducting each of the hybridizations together on the same surface,
as described above. Embodiments of the methods herein further include using any of the
calibration spots as described herein.

15   Another embodiment of the invention provided herein is a kit for analysis of genomic
abnormalities comprising a container and an array prepared by the method according to any of
methods herein. For example, the kit further comprises buffers for hybridization of the array
to a sample of nucleic acids. In a method of depositing a plurality of samples of a biological
material on a substrate in an array of elements having addressible locations, an improvement

20   is provided, the improvement comprising depositing the samples in a buffer having ionic
strength sufficient to produce a uniform distribution of the material throughout each element.
The biological material is typically DNA although the improvement is applicable to other
biological materials such as proteins. In particular with the arrays herein, the array includes
elements of genomic sequences of a plurality of chromosomes from cells from a plurality of

25   species of organism, and the sequences are syntenic.

       Embodiments of any of the kits herein can further include providing the kits with
arrays having a multi-array surface. Embodiments of the kits herein further include arrays
having elements with any of the calibration spots as described herein.

30   **Brief Description of the Drawings**

       Fig. 1a is a ratio plot of comparative genomic hybridization (CGH) of chromosome 18
of a 12 year old patient having delayed development and central nervous system
demyelination compared to a normal control, showing a deletion of 18 qter of approximately
7 megabases (Mb). The loss of genetic material is indicated by increase in Cy3™ labeled test

sample relative to Cy5™ labeled reference sample (indicated in red using the graphics software software commercially available with the arrays), at the right of the drawing which is the chromosome 18 q terminus.

Fig. 1b is a ratio plot of CGH of chromosome 4 of the same patient as in Fig. 1a, showing a gain in genetic material at 4q of 3.7 Mb. The gain in genetic material is indicated by the increase in Cy5™ labeled reference sample relative to Cy3™ labeled test sample (indicated in blue using the graphics software commercially available with the arrays) at the right of the drawing which is the chromosome 4 q terminus.

Fig. 2 is a ratio plot of the X chromosome of a male patient compared to a normal male control, showing a distal Xp duplication (at the p terminus of the X chromosome). The gain in genetic material is indicated by the increase in Cy5™ labeled reference sample relative to Cy3™ labeled test sample.

Detailed Description of Specific Embodiments

In array hybridization, an effective amount of genomic DNA obtained from cells of each of a test sample and a reference sample (e.g., a sample from cells known to be free of a chromosomal aberration) are each labeled with a detectable label, such as a fluorescent dye, and are each then hybridized to an array of nucleic acids obtained from each of a collection of BACs. Hybridization can be performed iteratively, on successive replicates of the array. The array contains cloned genomic DNA fragments that collectively cover substantially the entire genome of a chosen organism, such as a human. The resulting hybridization produces a fluorescently labeled array, the pattern of which reflects hybridization of sequences in the samples, i.e., the test genomic DNA and the reference genomic DNA, to homologous sequences within the arrayed BACs. For each test sample, a copy number, including possible deletions and insertions such as translocations, of every homologous sequence in each of the test and reference genomic DNA samples should directly affect the pattern of hybridization, both quantity and location, for example, as a fluorescent signal at discrete BACs located at known spots within the array. The versatility of the approach allows the detection of both constitutional variations in DNA copy number in clinical cytogenetic samples such as amniotic samples, chorionic villus samples (CVS), blood samples and tissue biopsies, as well as somatically acquired changes, for example, that arise during progression of cancers such as those in circulating blood cells, or in solid tumors.

The invention provides array-based methods, arrays and kits for determining genetic changes in a sample, such as a cell, a tissue or a cell culture population, compared to that in a

reference or normal sample. The methods and arrays of the invention provide greater levels of sensitivity, capable of detecting smaller genetic changes than previously available, and of detecting clonally distinct cell subpopulations. The methods and arrays of the invention are sufficiently sensitive to detect clonally distinct (by karyotypic criteria) cell populations within

5   a background cell population. Thus, the methods and arrays of the invention are particularly suited for accurate determination and analysis of the complex level of, for example, genetic mosaicism observed in many solid tumors and other tumorigenically altered cells and samples from individuals with an abnormal genetic make-up.

In one embodiment, the invention provides methods and an array for detecting genetic

10   mosaicism. Total genomic DNA is isolated from a cell population, e.g., a cancer cell population, with a known or unknown genetic constitution, for example, level of mosaicism. A predetermined level of genetic mosaicism can be obtained by conventional G-band karyotyping, also referred to as "GTG-banding technique" (see, e.g., Wakui 1999 J. Hum. Genet. 44:85-90); by fluorescence in situ hybridization ("FISH"; see, e.g., Zhao (2000)

15   Cancer Genet. Cytogenet. 118:108-111); or by spectral karyotyping ("SKY"; see, e.g., Veldman (1997) Nat. Genet. 15:406-410) or a combination thereof (see, e.g., Zhao (2001) Cancer Genet. Cytogenet. 127:143-147). Array-based genome profiling of the total genomic DNA from this cell population is obtained as described herein, and the number of clonal subpopulations with distinct karyotypes and their respective percentages in the total

20   population are measured. Data from the array-based profile are analyzed as a function of position on each chromosome of the test genome, and data from iterations of hybridization of both sample and reference nucleic acid, each labeled with each of at least a first and a second dye are compared, to determine precise chromosomal sites of a genetic abnormality.

In another embodiment, pre-isolated total genomic DNA from a homogenous

25   population of cells with a known genetic aberration or with a suspected genetic aberration is tested in comparison with isolated genomic DNA from cells having a "normal" or reference karyotype, e.g., cells with no known chromosomal aberrations. For example, the array genome profile on total genomic DNA has been established for a female abortus with a deletion of Xq and simultaneous trisomy of 16q. An effective amount of test genomic DNA

30   with normal 46,XX genomic DNA is used as a reference sample. Genetic aberrations detectable herein include those that are not visible by prior methods, i.e., those that may not be detected by conventional hybridization to an intact chromosomal metaphase spread.

By providing methods, apparatuses and kits having genomic arrays to determine the aberrant sites in a genome in a sample, i.e., genetic abnormalities, using the methods of the

invention, sites of such genetic abnormalities in each chromosome of a cell population can be accurately and efficiently determined.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

The terms "array" or "array" or "DNA array" or "nucleic acid array" or "chip" or "biochip" as used herein is a plurality of arrayed elements, each arrayed element comprising a defined amount of one or more species of biological molecules, e.g., a preparation of nucleic acids, immobilized on a substrate surface at a defined, i.e., at an addressible known location; as described in further detail, herein. In certain embodiments, an array of biological molecules may be an array of proteins (including peptides and polypeptides), carbohydrates, or lipids.

The term "aryl-substituted 4,4-difluoro-4-bora-3a, 4a-diaza-s-indacene dye" as used herein includes all "boron dipyrromethene difluoride fluorophore" or "BODIPY" dyes and "dipyrrometheneboron difluoride dyes" (see, e.g., U.S. Pat. No. 4,774,339), or equivalents, are a class of fluorescent dyes commonly used to label nucleic acids for their detection when used in hybridization reactions; see, e.g., Chen (2000) J. Org Chem. 65:2900-2906: Chen (2000) J. Biochem. Biophys. Methods 42:137-151. See also U.S. Pat. Nos. 6,060,324; 5,994,063; 5,614,386; 5,248,782; 5,227,487; 5,187,288.

The terms "cyanine 5" or "Cy5™" and "cyanine 3" or "Cy3™" refer to fluorescent cyanine dyes produced by Amersham Pharmacia Biotech (Piscataway, N.J.; Amersham Life Sciences, Arlington Heights, Ill.), as described in detail, herein, or equivalents. See U.S. Pat. Nos. 6,027,709; 5,714,386; 5,268,486; 5,151,507; 5,047,519. These dyes are typically incorporated into nucleic acids in the form of 5-amino-propargyl-2'-deoxy- cytidine 5'-triphosphate coupled to Cy5™ or Cy3™.

The terms "fluorescent dye" and "fluorescent label" as used herein includes all known fluors, including rhodamine dyes (e.g., tetramethylrhodamine, dibenzorhodamine, see, e.g., U.S. Pat. No. 6,051,719); fluorescein dyes; "BODIPY" dyes and equivalents (e.g., dipyrrometheneboron difluoride dyes, see, e.g., U.S. Pat. No. 5,274,113); derivatives of 1-[isoindolyl]methylene-isoindole (see, e.g., U.S. Pat. No. 5,433,896); and all equivalents. See also U.S. Pat. Nos. 6,028,190; 5,188,934.

The terms "hybridizing specifically to", "specific hybridization" and "selectively hybridize to," as used herein refer to formation of a nucleic acid base-paired duplex as a result of a high extent of complementary base pairing, of a target sample nucleic acid molecule or a target reference molecule, to a probe nucleotide sequence immobilized on a

5   substrate surface, under stringent conditions. The term "homologous" means that two nucleic acid sequences are sufficiently complementary by Watson-Crick rules of base pairing to hybridize under stringent conditions.

The term "stringent conditions" refers to conditions under which one nucleic acid of a given sequence will hybridize, i.e., will form a nucleic acid duplex preferentially with a

10   second nucleic acid sequence (e.g., a sample genomic nucleic acid hybridizing to an immobilized nucleic acid probe in an array), compared to forming a duplex to a lesser extent with, or not at all with, other sequences. A "stringent hybridization" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization (e.g., as in array, Southern or Northern hybridizations) are sequence dependent, and are different under

15   different environmental parameters. Generally, more stringent conditions are found at higher temperatures, and in the presence of agents that act to reduce the stability of hydrogen bonds, such as formamide. Stringent hybridization conditions as used herein can include, e.g., hybridization in a buffer comprising 50% formamide, 5 X SSC, and 1% SDS at 42°C, or hybridization in a buffer comprising 5 X SSC and 1% SDS at 65°C, both with a wash of 0.2

20   X SSC and 0.1% SDS at 65°C. Exemplary stringent hybridization conditions also include a hybridization buffer of 40% formamide, 1 M NaCl, and 1% SDS at 37°C, and a wash in 1 X SSC at 45°C. Those of ordinary skill will readily recognize that alternative but comparable hybridization and wash conditions can be utilized to provide conditions of similar stringency.

The precise hybridization format is not critical, since as is known in the art, it is

25   stringency of the wash conditions that determine whether a soluble, sample nucleic acid will specifically hybridize to an immobilized nucleic acid. Wash conditions can include, e.g.: a salt concentration of about 0.02 molar at pH 7 and a temperature of at least about 50°C or about 55°C to about 60°C; or, a salt concentration of about 0.15 M NaCl at 72°C for about 15 minutes; or, a salt concentration of about 0.2 X SSC at a temperature of at least about 50°C or

30   about 55°C to about 60°C for about 15 to about 20 minutes; or, the hybridization complex is washed twice with a solution with a salt concentration of about 2 X SSC containing 0.1% SDS at room temperature for 15 minutes and then washed twice by 0.1 X SSC containing 0.1% SDS at 68°C for 15 minutes; or, equivalent conditions. Stringent conditions for washing can also be, e.g., 0.2 X SSC/0.1% SDS at 42°C. See Sambrook, Ausubel, or Tijssen (cited

herein) for detailed descriptions of equivalent hybridization and wash conditions and for reagents and buffers, e.g., SSC buffers and equivalent reagents and conditions.

The term "karyotype" means the chromosomal aspect of the genome, or chromosome composition, of a cell or cell population. The term "karyotype" has also been used to mean

5   the appearance in a light microscope of a stained, complete chromosome set of the nucleus of a cell as the chromosomes appear during mitosis, and the chromosomal complement of an individual or sample, including the number of chromosomes and including any abnormalities which may be deviations from a normal or euploid set. In various embodiments, the methods of the invention can be used to determine deviations from the euploid set such as any

10   aneuploid variation in the karyotype of a cell population, whether the cell population is consistent in karyotype, or whether the cell population is characterized by genetic mosaicism, including the number of karyotype subpopulations in a sample and the percent of the cell population having a particular karyotype.

      Because specific inherited and acquired diseases and conditions have characteristic

15   karyotypes, determination of the chromosomally associated abnormalities of a cell or cell population can be used to diagnose, detect or prognose those diseases and conditions. Similarly, because levels of genetic abnormalities in a cancer or tumor population indicate a medical condition or a physiology, e.g., its tumorigenicity, determination of the karyotype of a cancer is useful for diagnosis, prognosis and treatment planning.

20   The phrase "labeled with a detectable composition" or " detectably labeled" as used herein refers to a nucleic acid comprising a detectable composition or moiety, i.e., a label, as described herein. The label can be another biological molecule, as a nucleic acid, e.g., a nucleic acid in the form of a stem-loop structure as a "molecular beacon," as described herein. The label can be colorimetrically or radioactively labeled bases (or, bases which can bind to a

25   detectable label), which can be incorporated into the nucleic acid by, e.g., nick translation, random primer extension, amplification with degenerate primers, and the like. The label can be detectable by any means, e.g., visual, spectroscopic, photochemical, bioluminescent, chemiluminescent, biochemical, fluorescent, immunochemical, physical or chemical means. Examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein

30   isothiocyanate (FITC), rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a chemiluminescent material is luminol; examples of bioluminescent materials are luciferase, luciferin, and aequorin.

The term "nucleic acid" as used herein refers to a deoxyribonucleotide or ribonucleotide in either single- or double-stranded form. The term encompasses nucleic acids

containing known analogues of natural nucleotides. The term also encompasses nucleic-acid-like structures with synthetic backbones. DNA backbone analogues provided by the invention include phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-

5   N-carbamate, morpholino carbamate, and peptide nucleic acids (PNAs); see Oligonucleotides and Analogues, a Practical Approach, edited by F. Eckstein, IRL Press at Oxford University Press (1991); Antisense Strategies, Annals of the New York Academy of Sciences, Volume 600, Eds. Baserga and Denhardt (NYAS 1992); Milligan (1993) J. Med. Chem. 36:1923-1937; Antisense Research and Applications (1993, CRC Press). PNAs contain peptide and

10   peptide-related backbones, such as N-(2-aminoethyl) glycine units. Phosphorothioate linkages are described, e.g., by U.S. Pat. Nos. 6,031,092; 6,001,982; 5,684,148; see also, WO 97/03211; WO 96/39154; Mata (1997) Toxicol. Appl. Pharmacol. 144:189-197. Other synthetic backbones encompassed by the term include methyl-phosphonate linkages or alternating methylphosphonate and phosphodiester linkages (see, e.g., U.S. Pat. No.

15   5,962,674; Strauss-Soukup (1997) Biochemistry 36:8692-8698), and benzylphosphonate linkages (see, e.g., U.S. Pat. No. 5,532,226; Samstag (1996) Antisense Nucleic Acid Drug Dev 6:153-156). The term "nucleic acid" as used herein structurally includes the terms gene, DNA, RNA, cDNA, mRNA, and chemically or enzymatically obtained derivatives and copies, including the terms oligonucleotide primer, probe and amplification product.

20   The term "genomic DNA" or "genomic nucleic acid" means nucleic acid isolated from a nucleus of one or more cells, and, includes nucleic acid derived from (e.g., isolated from, amplified from, cloned from, synthetic versions of) the total cellular or genomic DNA. The genomic DNA can be from any organismal source, including eukaryotic species, or from microorganisms which are prokaryotic, such as bacteria and blue-green algae, or acellular,

25   such as viruses.

          The term "a sample comprising a nucleic acid" or "sample of nucleic acid" as used herein refers to a sample comprising a DNA, an RNA, or nucleic acid representative of DNA or RNA isolated from a natural source, in a form suitable for hybridization (e.g., as a soluble aqueous solution) to another nucleic acid or combination thereof (e.g., hybrization to

30   immobilized probes or targets). The nucleic acid may be obtained as a plurality of isolated, cloned or amplified portions of a genome or gene; it may be, e.g., a genomic DNA, mRNA, or cDNA from substantially an entire genome, substantially all or part of a particular chromosome, or selected sequences (e.g. particular promoters, genes, amplification or restriction fragments, cDNA library, etc.). The nucleic acid sample may be extracted from

particular cells, tissues or body fluids, or, can be from cell cultures, including cell lines, or from preserved tissue sample, as described herein.

As used herein, the terms "computer" and "processor" are used in their broadest general contexts and incorporate all such devices. The methods of the invention can be

5   practiced using any computer/processor and in conjunction with any known software or methodology. For example, a computer/processor can be a conventional general-purpose digital computer, e.g., a personal "workstation" computer, including conventional elements such as microprocessor and data transfer bus. The computer/processor can further include any form of memory elements, such as dynamic random access memory, flash memory or the

10   like, or mass storage such as magnetic disc optional storage.

Generating and Manipulating Nucleic Acids

Practicing the methods of the invention involves isolation, synthesis, cloning, amplification, labeling and hybridization (e.g., hybridization) of nucleic acids. As described herein, nucleic acid for analysis and the immobilized nucleic acid on the array can be

15   representative of genomic DNA, including defined parts of, or entire, chromosomes, or entire genomes. Comparative genomic hybridization (hybridization) reactions, see, e.g., U.S. patent numbers 5,830,645, and 5,976,790. Nucleic acid samples are labeled with a detectable moiety, e.g., a fluorescent dye, for example, a first sample can labeled with a first dye and a second sample labeled with a second dye (e.g., Cy3™ and Cy5™). In one embodiment, the

20   test sample nucleic acid is labeled with at least one detectable moiety, e.g., a fluorescent dye that is different than is used to label a second or reference sample of nucleic acids, for example, for use in a first iteration of the hybridization. In a second iteration, the dyes used for labeling of the test sample and the reference sample are reversed, and data obtained from the first iteration is compared to that of the second iteration, as a control for any variables

25   such as efficiencies of labeling, detection of emission, and random non-specific binding.

In certain embodiments, the nucleic acids may be amplified using standard techniques such as PCR. Amplification can also be used to subclone or label the nucleic acid prior to the hybridization. The sample and/or the immobilized nucleic acid can be detectably labeled as described herein. The sample or the probe on the array can be produced from and collectively

30   can be representative of a source of nucleic acids from one or more particular (pre-selected) portions of a genome, e.g., a collection of polymerase chain reaction (PCR) amplification products, substantially an entire set of chromosomes, a selected chromosome or a chromosome fragment, or substantially an entire genome, e.g., as a collection of clones, e.g., BACs, PACs, YACs, and the like. The array-immobilized nucleic acid or genomic nucleic

acid sample may be processed in some manner prior to splitting or printing on the substrate, e.g., by blocking or removal of repetitive nucleic acids or by enrichment with selected nucleic acids.

5   Samples are applied to the immobilized probes (e.g., spotted or printed on the substrate to form the array) and, after hybridization and washing, the addressible location (e.g., spot on the array) and amount of each dye at each spot are read. The array-immobilized nucleic acid can be in the form of cloned DNA, e.g., YACs, BACs, PACs, and the like, as described herein. In one embodiment, each "spot" or probe element on the array has a known sequence, e.g., a known segment of the genome, and/or a known position on each of the

10   chromosomes of the genome, or other sequence.

General Techniques

Nucleic acids used to practice this invention, whether RNA, cDNA, genomic DNA, vectors, viruses or hybrids thereof, are isolated from any of a variety of sources, genetically engineered, amplified, and/or expressed/generated recombinantly. Any recombinant

15   expression system can be used, including, in addition to bacterial cells, e.g., mammalian, yeast, insect or plant cell expression systems.

Nucleic acids can be synthesized in vitro by well-known chemical synthesis techniques, as described in, e.g., Carruthers (1982) Cold Spring Harbor Symp. Quant. Biol. 47:411-418; Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res.

20   25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; U.S. Pat. No. 4,458,066. Double stranded DNA fragments may then be obtained either by chemically synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by

25   using the single strand as a template for enzymatically synthesizing the complementary strand using a DNA polymerase with a primer sequence.

Techniques for the manipulation of nucleic acids, such as, e.g., subcloning, labeling probes (e.g., random-primer labeling using Klenow polymerase, nick translation, amplification), sequencing, hybridization, G-banding, SKY, FISH and the like are well

30   described in the scientific and patent literature, see, e.g., Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN

BIOCHEMISTRY AND MOLECULAR BIOLOGY: hybridization WITH NUCLEIC ACID
PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Clones of Genomic Nucleic Acids

Genomic nucleic acids used in the methods, kits, and apparatus, of the invention, e.g.,
5   those immobilized onto arrays or used as samples, can be obtained and manipulated by
cloning into various vehicles. If necessary, genomic nucleic acid samples can be screened and
re-cloned or amplified from any source of genomic DNA. Thus, in various aspects, forms of
genomic nucleic acid used in the methods of the invention (including arrays and samples)
include genomic DNA, e.g., genomic libraries, which are contained in mammalian such as
10   human artificial chromosomes, satellite artificial chromosomes, yeast artificial chromosomes,
bacterial artificial chromosomes, P1 artificial chromosomes, and the like.

Mammalian artificial chromosomes (MACs) and human artificial chromosomes
(HAC) are, e.g., described in Ascenzioni (1997) Cancer Lett. 118:135-142; Kuroiwa (2000)
Nat. Biotechnol. 18:1086-1090; and U.S. patent numbers. 5,288,625; 5,721,118; 6,025,155;
15   and 6,077,697. MACs can contain inserts larger than 400 kilobase (Kb), see, e.g., Mejia
(2001) Am. J. Hum. Genet. 69:315-326. Auriche (2001) EMBO Rep. 2:102-107, has built
human minichromosomes having a size of 5.5 kilobase.

Satellite artificial chromosomes, or, satellite DNA-based artificial chromosomes
(SATACs), are, e.g., described in Warburton (1997) Nature 386:553-555; Roush (1997)
20   Science 276:38-39; Rosenfeld (1997) Nat. Genet. 15:333-335). SATACs can be made by
induced de novo chromosome formation in cells of different mammalian species; see, e.g.,
Hadlaczky (2001) Curr. Opin. Mol. Ther. 3:125-132; Csonka (2000) J. Cell Sci. 113 (Pt
18):3207-3216.

Yeast artificial chromosomes (YACs) can also be used and typically contain inserts
25   ranging in size from 80 to 700 kb. YACs have been used for stable propagation of genomic
fragments of up to one million base pairs in size; see, e.g., U.S. patent numbers 5,776,745;
and 5,981,175; Feingold (1990) Proc. Natl. Acad. Sci. USA 87:8637-8641; Tucker (1997)
Gene 199:25-30; Adam (1997) Plant J.11:1349-1358; and Zeschnigk (1999) Nucleic Acids
Res. 27:21.

30   Bacterial artificial chromosomes (BACs) are vectors that can contain inserts of size
120 Kb or greater, see, e.g., U.S. patent numbers 5,874,259; 6,277,621; and 6,183,957. BACs
are based on the E. coli F factor plasmid system, and DNA cloned into BACs can be purified
in microgram quantities. Because BAC plasmids are maintained in the bacterial cell at a copy

number of at one to two, unwanted genetic rearrangement observed with YACs are reduced
or eliminated; see e.g., Cao (1999) Genome Res. 9:763-774.

P1 artificial chromosomes (PACs), bacteriophage P1-derived vectors are described in
Woon (1998) Genomics 50:306-316; Reid (1997) Genomics 43:366-375; Nothwang (1997)

5   Genomics 41:370-378; and Kern (1997) Biotechniques 23:120-124). P1 is a bacteriophage
that infects E. coli that can contain 75 to 100 Kb DNA inserts. PACs are screened in much
the same way as lambda libraries.

Other cloning vehicles can also be used, for example, recombinant viruses; cosmids,
plasmids or cDNAs; see, e.g., U.S. patent numbers 5,501,979; 5,288,641; and 5,266,489.

10   These vectors can include marker genes, such as, e.g., luciferase and green fluorescent
protein genes (see, e.g., Baker (1997) Nucleic Acids Res 25:1950-1956). Sequences, inserts,
clones, vectors and the like can be isolated from natural sources, or can be obtained from
such sources as ATCC or GenBank libraries or commercial sources, or prepared by synthetic
or recombinant methods.

15   Amplification of Nucleic Acids

Amplification using oligonucleotide primers can be used to generate or manipulate,
e.g., subclone, genomic nucleic acids used in the methods of the invention, to incorporate
label into immobilized or sample nucleic acids, to detect or measure levels of nucleic acids
hybridized to an array, and the like. Amplification, typically with degenerate primers, is also

20   useful for incorporating detectable probes (e.g., Cy5™ or Cy3™-cytosine conjugates) into
nucleic acids representative of test or control genomic DNA to be used to hybridize to
immobilized genomic DNA. Amplification can be used to quantify the amount of nucleic
acid is in a sample, see, e.g., U.S. patent number 6,294,338. Amplification methods are also
well known in the art, and include, e.g., polymerase chain reaction (PCR; see, e.g., PCR

25   PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic
Press, N.Y. 1990, and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y.);
ligase chain reaction (LCR; see , e.g., Barringer 1990, Gene 89:117); transcription
amplification (see, e.g., Kwoh 1989, Proc. Natl. Acad. Sci. USA 86:1173); self-sustained
sequence replication (see, e.g., Guatelli 1990, Proc. Natl. Acad. Sci. USA 87:1874); Q Beta

30   replicase amplification (see, e.g., Smith 1997, J. Clin. Microbiol. 35:1477-1491); automated
Q-beta replicase amplification assay (see, e.g., Burg 1996, Mol. Cell. Probes 10:257-271);
and other RNA polymerase mediated techniques, e.g., nucleic acid sequence based
amplification, or, "NASBA" (see, e.g., Birch 2001, Lett. Appl. Microbiol. 33:296-301;
Greijer 2001, J. Virol. Methods 96:133-147).

Hybridizing Nucleic Acids

In practicing the methods of the invention, samples of nucleic acid, e.g., isolated, cloned or amplified genomic nucleic acid, are hybridized to immobilized nucleic acids. In various embodiments, the hybridization and/or wash conditions are carried out under

5   moderate to stringent conditions. An extensive guide to the hybridization of nucleic acids is found in, e.g., Sambrook Ausubel, Tijssen. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_M) for the specific sequence at a defined ionic strength and pH. The T_M is the temperature (under defined ionic strength and pH) at which 50% of the target reference sample sequence labeled

10   molecules hybridize to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_M for a particular probe. Exemplary stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on an array comprise 42°C using standard hybridization solutions (see, e.g., Sambrook), with the hybridization being carried out overnight. Exemplary highly stringent

15   wash conditions can comprise 0.15 M NaCl at 72°C for about 15 minutes. Exemplary stringent wash conditions can also comprise a 0.2 X SSC wash at 65°C for 15 minutes (see, e.g., Sambrook). In one aspect, a high stringency wash is preceded by a medium or low stringency wash to remove background probe signal. An exemplary medium stringency wash for a duplex of, e.g., more than 100 nucleotides, comprises 1 X SSC at 45°C for 15 minutes.

20   An exemplary low stringency wash for a duplex of, e.g., more than 100 nucleotides, can comprise 4 x to 6 X SSC at 40°C for 15 minutes.

In various embodiments, in practicing the array-based comparative hybridization (hybridization) reactions of the invention, the fluorescent dyes Cy3™. and Cy5™. are used to differentially label nucleic acid fragments from two samples, e.g., nucleic acid from a

25   reference or normal control is compared to a test sample nucleic acid from cell or tissue. Many commercial instruments are designed to accommodate to detection of two dyes. To increase the stability of Cy5™, or fluors or other oxidation-sensitive compounds, antioxidants and free radical scavengers can be used in hybridization mixes, the hybridization and/or the wash solutions. Thus, Cy5™ signals are dramatically increased and longer hybridization

30   times are possible. See co-pending U.S. patent application serial number 09/839,658, filed Apr. 19, 2001.

To further increase hybridization sensitivity, hybridization can be carried out in a controlled, unsaturated humidity environment; thus, hybridization efficiency is significantly improved if the humidity is not saturated. The hybridization efficiency can be improved if

the humidity is dynamically controlled, i.e., if the humidity changes during hybridization. Array devices comprising housings and controls that allow the operator to control the humidity during pre-hybridization, hybridization, wash and/or detection stages can be used. The device can have detection, control and memory components to allow pre-programming

5   of the humidity (and temperature and other parameters) during the entire procedural cycle, including pre-hybridization, hybridization, wash and detection steps.

The methods of the invention can incorporate hybridization conditions comprising temperature fluctuation. Hybridization has much better efficiency in a changing temperature environment as compared to conditions where the temperature is set precisely or at relatively

10   constant level (e.g., plus or minus a couple of degrees, as with most commercial ovens). Reaction chamber temperatures can be fluctuatingly modified by, e.g., an oven, or other device capable of creating changing temperatures.

The methods of the invention can comprise hybridization conditions comprising osmotic fluctuation. Hybridization efficiency (i.e., time to equilibrium) can also be enhanced

15   by a hybridization environment that comprises changing hyper-/hypo-tonicity, e.g., a solute gradient. A solute gradient is created in the device. For example, a low salt hybridization solution is placed on one side of the array hybridization chamber and a higher salt buffer is placed on the other side to generate a solute gradient in the chamber.

Fragmentation and Digestion of Nucleic Acid

20   In practicing the methods of the invention, immobilized and sample nucleic acids can be cloned, labeled or immobilized in a variety of lengths. For example, in one aspect, the genomic nucleic acid can have a length smaller than about 200 bases. Use of labeled genomic DNA limited to this small size significantly improves the resolution of the molecular profile analysis, e.g., in array-based hybridization. For example, use of such small fragments allows

25   for significant suppression of repetitive sequences and other unwanted, "background" cross-hybridization on the immobilized nucleic acid. Suppression of repetitive sequence hybridization greatly increases the reliability of the detection of copy number differences (e.g., amplifications or deletions) or detection of unique sequences.

The resultant fragment lengths can be modified by, e.g., treatment with DNase.

30   Adjusting the ratio of DNase to DNA polymerase in a nick translation reaction changes the length of the digestion product. Standard nick translation kits typically generate 300 to 600 base pair fragments. If desired, the labeled nucleic acid can be further fragmented to segments of 200 bases, down to as low as about 25 to 30 bases. Random enzymatic digestion of the DNA is carried out, using, e.g., a DNA endonuclease, e.g., DNase (see, e.g., Herrera

(1994) J. Mol. Biol. 236:405-411; Suck (1994) J. Mol. Recognit. 7:65-70), or, the two-base restriction endonuclease CviJI (see, e.g., Fitzgerald (1992) Nucleic Acids Res. 20:3753-3762) and standard protocols, see, e.g., Sambrook, Ausubel, with or without other fragmentation procedures.

5   Other procedures can also be used to fragment genomic DNA, e.g. mechanical shearing, sonication (see, e.g., Deininger (1983) Anal. Biochem. 129:216-223), and the like (see, e.g., Sambrook, Ausubel, Tijssen). For example, one mechanical technique is based on point-sink hydrodynamics that result when a DNA sample is forced through a small hole by a syringe pump, see, e.g., Thorstenson (1998) Genome Res. 8:848-855. Fragment size can be

10   evaluated by a variety of techniques, including, e.g., sizing electrophoresis, as by Siles (1997) J. Chromatogr. A. 771:319-329, that shows analysis of DNA fragmentation using a dynamic size-sieving polymer solution in a capillary electrophoresis. Fragment sizes can also be determined by, e.g., matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, see, e.g., Chiu (2000) Nucleic Acids Res. 28:E31.

15   Comparative Genomic Hybridization (hybridization)

           The methods of the invention are used in array-based comparative genomic hybridization reactions to detect a chromosomal abnormality in the sample, or detect genetic mosaicism in cell populations, such as tissue, e.g., biopsy or body fluid samples. Hybridization is a molecular cytogenetics approach that can be used to detect regions in a

20   genome undergoing changes, e.g., gains or losses of a sequence or of a change in copy numbers of a sequence. Analysis of genomes of tumor cells can detect a region or regions of anomaly and to monitor the ongoing process.

           Hybridization reactions compare the genetic composition of test versus controls samples; e.g., whether a test sample of genomic DNA (e.g., from a cell population suspected

25   of having one or more genetic defects) has amplified or deleted or mutated segments, as compared to a reference sample which is a "negative" control, e.g., "normal" or "wild type" genotype, or to a "positive" control, e.g., a known cancer cell or a cell with a known defect, e.g., a known translocation or deletion or amplification or the like.

           The methods of the invention can be practiced with all known methods and means and

30   variations thereof for carrying out comparative genomic hybridization, see, e.g., U.S. patent numbers 6,197,501; 6,159,685; 5,976,790; 5,965,362; 5,856,097; 5,830,645; 5,721,098; 5,665,549; and 5,635,351; and, Diago (2001) American J. of Pathol. May;158(5):1623-1631;(Theillet (2001) Bull. Cancer 88:261-268; Werner (2001) Pharmacogenomics 2:25-36; Jain (2000) Pharmacogenomics 1:289-307.

Arrays, or "BioChips"

The present invention in one embodiment provides arrays and methods of producing them. In an alternative embodiment, the methods herein can be practiced with any known "array," also referred to as a "array" or "DNA array" or "nucleic acid array" or "biochip," or

5   variation thereof. An array is generically a plurality of "probe elements," or "spots," each probe element comprising a defined amount of one or more biological molecules, e.g., polypeptides, nucleic acid molecules, or probes, immobilized on a known or defined (addressible) location on a substrate surface. Typically, the immobilized biological molecules are contacted with a sample for specific binding, e.g., hybridization, between molecules in the

10   sample and the array. Immobilized nucleic acids can contain sequences from specific messages (e.g., as cDNA libraries) or genes (e.g., genomic libraries), including, e.g., substantially all or a subsection of a chromosome or substantially all of a genome, including a human genome. Other elements or spots can contain reference sequences, such as positive and negative controls, and the like. The elements of the arrays may be arranged on the

15   substrate surface at different sizes and different densities. Different probe elements of the arrays can have the same molecular species, e.g., in different amounts, densities, sizes, labeled or unlabeled, and the like.

The sizes and densities of the spots or elements sizes depend upon a number of factors, such as the nature of the label, the substrate support (which is solid, semi-solid,

20   fibrous, capillary or porous), and the like. Each spot or element may comprise substantially the same nucleic acid sequences, or alternatively can be a mixture of nucleic acids of different lengths and/or sequences. Thus, for example, a spot or element contains more than one copy of a cloned piece of DNA, and each copy may be broken into fragments of different lengths, by methods described herein.

25   In one embodiment, the array can contain spots encoding all or substantially all of syntenic sequences from highly homologous sequences of chromosomes of two or more different species. Syntenic sequences are those that are located on the same chromosome, for example, a large chromosome such as human chromosome 1, or a small chromosome such as human X chromosome, whether or not these are linked by classical genetic analysis. In one

30   embodiment, a syntenic an array contains elements having syntenic sequences from one or more chromosomes that contain homologous sequences from each of a human and another vertebrate, such as a mammal, e.g., a human and a rodent such as a mouse, or a human and a chimpanzee. Alternatively, the vertebrate can be a non-mammal, such as a frog (*Rana*) or an African clawed toad (*Xenopus*) or a zebra fish. In another embodiment, a syntenic array

23

contains elements having syntenic sequences from one or more chromosomes from a human and an invertebrate species of interest, for example, a human and a *Drosophila*. The immobilized sequences, obtained for example from cloned BAC libraries using clones known to be syntenic from each species, are placed in the array at various addressible locations.

5   A syntenic array having arrays of spots of nucleic acids with, for example, identifiable regions of chromosomes from a human and from another species such as a mouse can be prepared using libraries such as BAC libraries comprising the entire genome, that are known to one of skill in the art. These can then be used to analyze samples obtained from human subjects having diseases or disease conditions, and from animal subjects for which animal

10   models are known that are models for the human disease. Examples of diseases for which animal models are known include lung cancer, mesothelioma, adenocarcinoma, and prostate cancer.

Alternatively, these arrays can be used to test genotoxicity, i.e., mutagenicity, of chemical compositions towards genomes of test animals, by analysis of the genome of cells,

15   organisms, or cell populations exposed to or admininstered the chemical composition. Cell populations can be obtained from organisms exposed to the chemical, or can be cultured in vitro or ex vivo and then exposed to the chemical composition. The arrays of the invention, by testing the effect of the chemical on the entire genome, can replace or supplement present toxological analyses that require larger numbers of organisms or longer periods of analysis,

20   while yielding results showing effect of the chemical on the entire genome, and relating those analyses to the human genome.

Syntenic arrays can be prepared having homologous sequences in a plurality of organisms, such that samples taken from mouse cells exposed to a chemical, or from mouse cells having a particular cancer, can, for example, be analyzed with syntenic immobilized

25   arrayed probe sequences from a human, a mouse, a dog, and a frog, simultaneously and in a time efficient manner.

The array can comprise nucleic acids immobilized on any substrate, e.g., a solid surface (e.g., nitrocellulose, glass, quartz, fused silica, plastics and the like). See, e.g., U.S. patent number 6,063,338 describing multi-well platforms comprising cycloolefin polymers if

30   fluorescence is to be measured. Arrays used in the methods of the invention can include housing having components for controlling humidity and temperature during the hybridization and wash reactions.

In practicing the methods of the invention, known arrays and methods of making and using arrays can be incorporated in whole or in part, or variations thereof, as described, for

24

example, in U.S. patent numbers 6,277,628; 6,277,489; 6,261,776; 6,258,606; 6,054,270; 6,048,695; 6,045,996; 6,022,963; 6,013,440; 5,965,452; 5,959,098; 5,856,174; 5,830,645; 5,770,456; 5,632,957; 5,556,752; 5,143,854; 5,807,522; 5,800,992; 5,744,305; 5,700,637; 5,556,752; and 5,434,049; see also, e.g., WO 99/51773; WO 99/09217; WO 97/46313; WO

5   96/17958; see also, e.g., Johnston (1998) Curr. Biol. 8:R171-R174; Schummer (1997) Biotechniques 23:1087-1092; Kern (1997) Biotechniques 23:120-124; Solinas-Toldo (1997) Genes, Chromosomes & Cancer 20:399-407; Bowtell (1999) Nature Genetics Supp. 21:25-32. See also published U.S. patent applications Nos. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765. The present

10   invention in various embodiments can use any known array, e.g., GeneChips™, Affymetrix, Santa Clara, CA; SpectralChip™ Mouse BAC Arrays, SpectralChip™ Human BAC Arrays and Custom Arrays of Spectral Genomics, Houston, Texas, and their accompanying manufacturer's instructions.

Substrate Surfaces

15   The arrays used to practice the invention can have substrate surfaces of a rigid, semi-rigid or flexible material. The substrate surface can be flat or planar, be shaped as wells, raised regions, etched trenches, pores, beads, filaments, or the like. Substrates can be of any material upon which a "capture probe" can be directly or indirectly bound. For example, suitable materials can include paper, glass (see, e.g., US patent number 5,843,767), ceramics,

20   quartz or other crystalline substrates (e.g. gallium arsenide), metals, metalloids, polacryloylmorpholide, various plastics and plastic copolymers, Nylon.TM., Teflon.TM., polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polystyrene/latex, polymethacrylate, poly(ethylene terephthalate), rayon, nylon, poly(vinyl butyrate), polyvinylidene difluoride (PVDF); (see, e.g., U.S. patent number 6,024,872), silicones (see,

25   e.g., U.S. patent number 6,096,817), polyformaldehyde (see, e.g., U.S patent numbers 4,355,153; and 4,652,613), cellulose (see, e.g., U.S. patent number 5,068,269), cellulose acetate (see, e.g., U.S. patent number 6,048,457), nitrocellulose, various membranes and gels (e.g., silica aerogels, see, e.g., U.S. patent number 5,795,557), paramagnetic or superparamagnetic microparticles (see, e.g., U.S. patent number 5,939,261) and the like.

30   Reactive functional groups can be, e.g., hydroxyl, carboxyl, amino groups or the like. Silane (e.g., mono- and dihydroxyalkylsilanes, aminoalkyltrialkoxysilanes, 3-aminopropyl-triethoxysilane, 3-aminopropyltrimethoxysilane) can provide a hydroxyl functional group for reaction with an amine functional group.

Nucleic Acids and Detectable Moieties: Incorporating Labels and Scanning Arrays

25

The methods of the invention use nucleic acids associated with a detectable label, e.g., have incorporated or have been conjugated to a detectable moiety. Any detectable moiety can be used. The association with the detectable moiety can be covalent or non-covalent. In another embodiment, the array-immobilized nucleic acids and sample nucleic acids are

5   differentially detectable, e.g., they have different labels and emit difference signals. In yet another embodiment, the array-immobilized nucleic acids are unloaded, and the test sample nucleic acid and the reference nucleic acid are differentially blended, and in another iteration of the hybridization, the differential labels are changed.

Useful detectible labels or tags include, e.g., radioactive labels, e.g., ^{32}P, ^{35}S, ^{3}H, ^{14}C,

10   ^{125}I, ^{131}I; fluorescent dyes (e.g., Cy5™, Cy3™, FITC, rhodamine, lanthanide phosphors, Texas red), electron-dense reagents (e.g. gold), enzymes, e.g., as commonly used in an ELISA (e.g., horseradish peroxidase, β-galactosidase, luciferase, alkaline phosphatase), colorimetric labels (e.g. colloidal gold), magnetic labels (e.g. Dynabeads™), biotin, dioxigenin, or any haptens and proteins for which antisera or monoclonal antibodies are

15   available. The label can be directly incorporated into the nucleic acid to be detected, or it can be attached to a probe or antibody having a linked moiety that hybridizes or binds to the nucleic acid. A peptide can be made detectable by incorporating (e.g., into a nucleoside base) a predetermined polypeptide epitope recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, transcriptional activator polypeptide,

20   metal binding domains, epitope tags). Label can be attached by spacer arms of various lengths to reduce potential steric hindrance or impact on other useful or desired properties. See, e.g., Mansfield (1995) Mol Cell Probes 9:145-156. In array-based hybridization, fluors can be paired together; for example, one fluor labeling the control sample (e.g., the nucleic acid of known, or normal, karyotype) and another fluor the test sample nucleic acid (e.g.,

25   from a CVS or from a cancer cell sample). Exemplary pairs are: rhodamine and fluorescein (see, e.g., DeRisi (1996) Nature Genetics 14:458-460); lissamine-conjugated nucleic acid analogs and fluorescein-conjugated nucleotide analogs (see, e.g., Shalon (1996) supra); Spectrum Red.TM. and Spectrum Green.TM. (Vysis, Downers Grove, Ill.); Cy3™ and Cy5™ Cy3 ™ and Cy5™ can be used together; both are fluorescent cyanine dyes produced by

30   Amersham Life Sciences (Arlington Heights, Ill.). Cyanine and related dyes, such as merocyanine, styryl and oxonol dyes, are particularly strongly light-absorbing and highly luminescent, see, e.g., U.S. patent numbers 4,337,063; 4,404,289; and 6,048,982.

Other fluorescent nucleotide analogs can be used, see, e.g., Jameson (1997) Methods Enzymol. 278:363-390; Zhu (1994) Nucleic Acids Res. 22:3418-3422. U.S. patent numbers

5,652,099 and 6,268,132 also describe nucleoside analogs for incorporation into nucleic acids, e.g., DNA and/or RNA, or oligonucleotides, via either enzymatic or chemical synthesis to produce fluorescent oligonucleotides. U.S. patent number 5,135,717 describes phthalocyanine and tetrabenztriazaporphyrin reagents for use as fluorescent labels.

5   Detectable moieties can be incorporated into genomic nucleic acid of the test or reference sample, and, if desired, into "target" nucleic acid, by covalent or non-covalent means, e.g., by transcription, such as by random-primer labeling using Klenow polymerase, or "nick translation," or, amplification, or equivalent. For example, in one aspect, a nucleoside base is conjugated to a detectable moiety, such as a fluorescent dye, e.g., Cy3™ or

10   Cy5™, and then incorporated into a sample genomic nucleic acid. Samples of genomic DNA can be incorporated with a Cy3™- or a Cy5™-dCTP conjugate mixed with unlabeled dCTP. Cy5™ is typically excited by the 633 nm line of HeNe laser, and emission is collected at 680 nm. See also, e.g., Bartosiewicz (2000) Archives of Biochem. Biophysics 376:66-73; Schena (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Pinkel (1998) Nature Genetics 20:207-

15   211; Pollack (1999) Nature Genetics 23:41-46.

In another embodiment, for performing PCR or nick translation to label nucleic acids, modified nucleotides synthesized by coupling allylamine-dUTP to the succinimidyl-ester derivatives of the fluorescent dyes or haptenes (such as biotin or digoxigenin) are used; this method allows custom preparation of most common fluorescent nucleotides, see, e.g.,

20   Henegariu (2000) Nat. Biotechnol. 18:345-348.

In certain embodiments of the methods of the invention, labeling with a detectable composition (labeling with a detectable moiety) also can include a nucleic acid attached to another biological molecule, such as a nucleic acid, e.g., a nucleic acid in the form of a stem-loop structure as a "molecular beacon" or an "aptamer beacon." Molecular beacons as

25   detectable moieties are well known in the art; for example, Sokol (1998) Proc. Natl. Acad. Sci. USA 95:11538-11543, synthesized "molecular beacon" reporter oligodeoxynucleotides with matched fluorescent donor and acceptor chromophores on their 5' and 3' ends. In the absence of a complementary nucleic acid strand, the molecular beacon remains in a stem-loop conformation where fluorescence resonance energy transfer prevents signal emission. On

30   hybridization with a complementary sequence, the stem-loop structure opens increasing the physical distance between the donor and acceptor moieties thereby reducing fluorescence resonance energy transfer and allowing a detectable signal to be emitted when the beacon is excited by light of the appropriate wavelength. See also, e.g., Antony (2001) Biochemistry

40:9387-9395, describing a molecular beacon comprised of a G-rich 18-mer triplex forming oligodeoxyribonucleotide. See also U.S. patent numbers 6,277,581 and 6,235,504.

Aptamer beacons are similar to molecular beacons; see, e.g., Hamaguchi (2001) Anal. Biochem. 294:126-131; Poddar (2001) Mol. Cell. Probes 15:161-167; Kaboev (2000) Nucleic

5   Acids Res. 28:E94. Aptamer beacons can adopt two or more conformations, one of which allows ligand binding. A fluorescence-quenching pair is used to report changes in conformation induced by ligand binding. See also, e.g., Yamamoto (2000) Genes Cells 5:389-396; and Smirnov (2000) Biochemistry 39:1462-1468.

In addition to methods for labeling nucleic acids with fluorescent dyes, methods for

10   the simultaneous detection of multiple fluorophores are well known in the art, see, e.g., U.S. patent numbers 5,539,517; 6,049,380; 6,054,279; and 6,055,325. For example a spectrograph can image an emission spectrum onto a two-dimensional array of light detectors; a full spectrally resolved image of the array is thus obtained. Photophysics of the fluorophore, e.g., fluorescence quantum yield and photodestruction yield, and the sensitivity of the detector are

15   read time parameters for an oligonucleotide array. With sufficient laser power and use of Cy5™ and/or Cy3™, which have lower photodestruction yields, an array can be read in less than 5 seconds.

It is desirable for detection and analysis of a mixture of two or more fluors or fluorescent dyes such as Cy3™ and Cy5™, to create a composite image showing the amount

20   of each of the plurality of fluors. To acquire the two or more images, the array can be scanned either simultaneously or sequentially. Charge-coupled devices, or CCDs, are used in array scanning systems, including practicing the methods of the invention. Thus, CCDs used in the methods of the invention can scan and analyze multicolor fluorescence images.

Devices and methods can be used or adapted to practice the methods of the invention,

25   including array reading or "scanning" devices, for scanning the array following hybridization, and for further analyzing multicolor fluorescence images; see, e.g., U.S. patent numbers 6,294,331; 6,261,776; 6,252,664; 6,191,425; 6,143,495; 6,140,044; 6,066,459; 5,943,129; 5,922,617; 5,880,473; 5,846,708; and 5,790,727; and, the patents cited in the discussion of arrays herein. See also published U.S. patent applications having numbers 20010018514; and

30   20010007747.

The methods of the invention further comprise data analysis, which can include the steps of determining, e.g., fluorescent intensity as a function of substrate position, removing "outliers" (data deviating from a predetermined statistical distribution), or calculating the relative binding affinity of the immobilized array targets from the remaining data. The

resulting data can be displayed as an image with color in each region varying according to the light emission or binding affinity between targets and probes. See, e.g., U.S. patent numbers 5,324,633; 5,863,504; and 6,045,996. The invention can also incorporate a device for detecting a labeled marker on a sample located on a support, see, e.g., U.S. patent number

5   5,578,832.

In one embodiment, the data are displayed as a ratio plot of normalized data from two independent arrays, for example, in which Cy3™ -labeled test sample data normalized to Cy5™-labeled reference sample are shown in red, and Cy5™-labeled test sample data normalized to Cy3™-labeled reference data, are shown in blue. The normalized ratio,

10   displayed on the ordinate, from each of the individual clones, is displayed linerearly ordered according to position on a chromosome along the abscissa, such that the p-arm terminus clone is displayed on the left, and the q-arm terminus is displayed on the right, for each chromosome. Reciprocal values (each normalized to a reference control) are used for red and blue plots, so that the obtained red and blue functions deviate in opposite directions (one

15   positive and one negative) if a genetic abnormality that is a deletion of one or more BAC clones, is significant. Similarly, the two functions deviate in the same direction (both positive) if a genetic abnormality that is an insertion compared to one or more BAC clones, is observed. Non-significant deviations that are due, e.g., to non-specific binding or other random effects, appear in only one of the two functions.

20   Sources of Genomic Nucleic Acid for Sample Preparation

The invention provides methods of detecting a chromosomal abnormality in a sample comprising nucleic acid, such as a cell population or a tissue or fluid sample, by performing an array-based comparative genomic hybridization. The nucleic acid can be isolated from or, amplified from or, cloned from genomic DNA. The genomic DNA can be from any source,

25   for example, the cell, tissue or fluid sample from which the nucleic acid sample is prepared is taken from a subject or a cell exposed to a chemical composition or a physical force, to determine whether the composition or force is associated with a genetic defect, and to compare the abnormality with that of a patient having or suspected of having a pathology or a condition. A causality relationship may be established between the composition or force and

30   the chromosomal abnormality, or the diagnosis or prognosis of the pathology or condition can be associated with a genetic defect, e.g., a cancer or tumor comprising cells with genomic nucleic acid base substitutions, amplifications, deletions and/or translocations. The test sample (and a control reference sample) can be a cell sample, such as tissue or fluid from, e.g., amniotic samples, CVS, serum, blood, chord, blood or urine samples, central nervous

system (CNS) or bone marrow aspirations, fecal samples, saliva, tears, tissue and surgical biopsies, needle or nucleic acids from punch biopsies, and the like. The reference sample may be a standard that is used for each analysis, i.e., a uniform standard as a negative control for chromosomal abnormalities, or a positive control for a particular known syndrome, defect, or

5   disease.

Methods of isolating cell, tissue or fluid samples are well known to those of skill in the art sample sources include, but are not limited to, aspirations, tissue sections, drawing of blood or other biological fluids, surgical or needle biopsies, and the like. A "clinical sample" derived from a patient includes frozen sections or paraffin sections taken for histological

10   purposes. The sample can also be derived from supernatants (of cell cultures), lysates of cells, cells from tissue culture in which it may be desirable to detect a genetic abnormality, including chromosomal abnormalities and changes in gene or chromosomal copy numbers.

Chromosomal Abnormalities

The methods, arrays and kits of the invention can be used for detecting genotoxic

15   effects of a chemical composition or preparation, or a physical force, and can also be used for diagnosing diseases and conditions, formulating appropriate treatment plans and preparing a prognosis, using one or more models from a heterologous species, to extrapolate to the species of interest, i.e., a human. The methods and arrays herein provide a surrogate for use of human cells, by analyzing syntenic strands of chromosomes from more than one species or

20   genus on a single surface. The syntenic arrays can be arranged, by selection of clones that carry nucleic acid sequences from numbers and locations at selected portions of an organisms' genome, to focus on chromosomes or parts of chromosomes of interest in two or more species of organism, or can include an entire genome of at least one of a plurality of species of organisms. Causality of chromosome defects or abnormalities can be associated

25   with one or more compositions or forces, and further associated with one or more known genetic defects. Further, methods, apparatuses and kits of the invention can be used for analyzing progression of a disease, e.g., a cancer or tumor comprising cells with genomic nucleic acid base substitutions, amplifications, deletions and/or translocations, or, an inherited condition. In some situations, the amount or degree of different subpopulations

30   comprising different genetic makeups (karyotypes) in a tumor or other cancer cell population from a patient can be helpful in classifying the cancer or formulating a treatment plan or prognosis.

Chromosome abnormalities are also common causes of congenital malformations and spontaneous abortions. They include structural abnormalities such as deletions, insertions,

translocations and amplifications of various portions of chromosome; polyploidy; monosomy; trisomy; and mosaicism.

Methods of the invention can also be used to detect aneuploidy (a deviation from euploidy, commonly complete diploidy) of all or parts of one or more chromosomes, for

5   example, chromosomes 13, 18, 21, X, and Y from genomic DNA from newborn uncultured blood samples (see, e.g., Jalal (1997) Mayo Clin. Proc. 72:705-710). Mosaicism has been found in approximately 1%-2% of viable pregnancies as determined by CVS at 9-11 weeks of gestation. It has been detected in pregnancies with both diploid and trisomic fetuses and appears to have an important effect on the intrauterine fetal survival, see, e.g., Harrison

10 (1993) Hum. Genet. 92:353-358. Experimental cells of human origin or from another species are used to analyze for aneuploidy produced by a chemical composition or a physical force, and arrays are designed to report abnormalities along a syntenic set of clones of a chromosome and its homolog in another species.

Preimplantation genetic diagnosis of oocytes and embryos has become the technique

15 of choice to select against abnormal embryos before embryo transfer in vitro fertilization (IVF) programs. Thus, in another embodiment, the methods of the invention are used for preimplantation genetic diagnosis and the diagnosis of structural abnormalities in oocytes and embryos. See, e.g., Fung (2001) J. Histochem. Cytochem. 49:797-798. The methods, apparatuses and kits of the invention are useful in conjunction with CVS and fetal

20 karyotyping. See, e.g., Sanz (2001) Fetal Diagn. Ther. 16:95-97.

Genetic mosaicism is frequent among transgenic animals produced by pronuclear microinjection. A successful method for the screening of founder animals for germline mosaicism prior to mating would greatly reduce the costs associated with the propagation of the transgenic lines, and improve the efficiency of transgenic livestock production. A

25 syntenic array using the methods herein enables analysis of animals of a variety of species on a single surface, such as a glass slide. In each analysis, two mixtures of test and reference nucleic acids are made: the test sample separately labeled with a each of a first detectable label and a different and distinguishable second is mixed with a reference sample labeled with each of the second detectable label and the first detectable label, respectively. The two

30 mixtures are hybridized to iterations of the syntenic array on a surface having a plurality of arrays, i.e., on a multi-array surface. Each of the iterations of the array has elements having nucleic acids from the syntenic strand of the chromosome from each of the species. Results obtained for each species are compared directly for results obtained from a second or more species. Thereafter, one of the species can be used as a surrogate for analysis of the chemical

31

composition, physical force, or progression of the condition. The methods of the invention are useful in the production of transgenic animals, particularly, the screening of founder animals for germline mosaicism prior to mating. See, e.g., Ibanez (2001) Mol. Reprod. Dev. 58:166-172.

5   Toxicology Using Whole Organisms and Cells in Cultures

Methods herein are suitable for analysis of abnormalities in cell samples obtained from organisms (species of animals, plants, fungi, etc.) and from samples of cells in culture. Cells may be primary cultures or established cell lines. Primary cultures may be obtained cell samples taken from whole metazoan animals or from multi-cellular plants, for example, that

10   have been exposed (test samples) to a chemical or physical agent in an environment, or from unexposed control organisms which may be used as reference samples. Organisms, primary cultures, and established cell lines are exposed to an agent under rigorous controlled laboratory conditions, or are placed in a natural environment to monitor that environment for the presence of a deleterious agent, or are feral organism obtained from the environment as a

15   means of monitoring a possible past history of genotoxic agents in that environment.

Primary cell cultures are obtained from tissues and organs by culture methods using media known to one of ordinary skill in the art of cell cultures. Suitable animal cell types are ascites, epithelial (including epidermal, tracheal/bronchial, renal tubule, hematopoietic) endothelial, meural cells, lymphocytes and the like. Cells may be obtained from wild-type

20   outbred strains such as BN (Brown Norway) rat, or from inbred isogenic strains such as F344 rat, some of which exhibit substantially greater sensitivity to chemical agents than outbred strains, for example, are tens to hundreds fold more sensitive as indicated by presenting similar effects at tens to hundreds fold decreased dose of the agent. Suitable rodent strains are obtained from Charles River Laboratory, Wilmington, MA or other suppliers.

25   Toxicology studies using the syntenic arrays herein are also performed with genetically engineered organisms or with mutant strains, any of which may be heterozygous or homozygous for one or more transgenes or mutant alleles (see Dashwood, R., J. Biochem. Mol. Biol. 36(1)35-42, 2003). Mutant animals such as the non-obese diabetic mouse (NOD), or animals treated to obtain a model system of a human disease, such as streptozotocin-

30   treated diabetic mice, or myelin-treated mice having experimental allergic encephalomyelitis (EAE, a model for multiple sclerosis) may be used. Further, mouse model strains carrying engineered reporter genes such as *lacZ, lacI,* c-*myc/lacZ, rpsL,* and *gptΔ* transgenic animals, and knock-out strains such as $p53^{+/-}$, $XPA^{-/-}$, $XPL^{-/-}$, and the like have been used to detect frequency and nature of point mutations or deletions spanning one or a few base pairs within

the gene. Mouse mutant strains are available from Jackson Laboratories, Bar Harbor, ME and other suppliers. While none of these strains, however, is engineered to be capable of detecting large chromosomal abnormalities such as deletions, amplifications, or other major chromosomal changes in response to an agent in an environment, any of these cell lines or

5   strains, or rodents in vivo, can be analyzed by the methods herein for such chromosomal abnormalities. DNA is prepared from a cell sample and labelled according to Example 2 herein, and tested on the syntenic arrays designed for use both with human and non-human species, e.g., rodent species such as rat or mouse, as described in Example 4. Data obtained is complementary to that obtained by the use of such reporter genes, and both data can be

10   obtained from the same groups of animals.

Advantageously, data obtained using the syntenic arrays herein indicate both the presence of a chromosomal abnormality in the test samples, and its location on the chromosome of the test organism, e.g., rat, mouse, guinea pig, further indicates its location on a homologous chromosome of another species, such as a human. The test organism therefore

15   is used herein as a surrogate as prospective designed toxicological tests on humans cannot be performed.

Multi-array Surfaces

Multi-array surfaces provided herein have on each surface a plurality of copies of the array, i.e., arrays of biological molecules, for example, nucleic acids. The term "multi-array

20   surface" or "surface" as used herein means an article of manufacture having a plurality of micro-arrays applied to a side or a face of a substrate. In general the micro-arrays are printed or spotted or otherwise deposited on the face of the substrate, in an arrangement such that the arrays are non-contiguous, i.e., the arrays are distal from each other on the surface, or are not in contact, compared to the size of each array and the spacings of the spots within each array.

25   A multi-array surface having a plurality of arrays is desirable for the following procedures: hybridizations are conducted in duplicate or triplicate on a single surface. Previous to the present invention, duplicate or triplicate or even a greater number of replicated spots have been described that are present on a single surface, however all spots were exposed to hybridization of a single hybridization mixture prepared from a biological

30   sample. The hybridization mixture is a solution that typically contains a nucleic acid sample from a test subject or a reference subject, and is labeled with a fluorescent dye, or is a mixture of two different samples of nucleic acids of different origins, each labeled with the same or a different dye. The hybridization mixture is formed prior to hybridization with the spots or elements of the array on the surface, for example, the mixture includes nucleic acids from test

subject labeled with a first fluorescent dye and nucleic acids from a reference sample labeled with a different and distinguishable dye. The reference sample can be nucleic acids from a normal individual of the same species as the test subject, or can be nucleic acids of a different species, or nucleic acids from a single BAC clone or from a mixture of BAC clones. For

5   BAC clones, NCBI maintains a human BAC resource, which provides genome-wide information concerning large-insert clones that integrate cytogenetic, radiation-hybrid, linkage, and sequence maps of the genome. See www.ncbi.nlm.nih.gov/genome/cyto/hrbc.shtml.

It is desirable, in analyzing such data, to perform the hybridization in two different

10   formats that reverse the fluorescent labels, what is commonly described as a "label reversal", "label swap" or "dye swap" analysis. In a dye swap analysis, at least two nucleic acid samples are to be compared, and at least two mixtures are made. In the first mixture, a first label such as a first fluorescent dye is used to identify the reference nucleic acid probe, and a second label such as a second fluorescent dye is used to identify the test sample, and after

15   labeling each, the mixture is made. Then the labels are reversed, i.e., a second mixture is made in which the reference nucleic acid probe carries the second dye and the test sample carries the first dye. Each of the two mixtures provides a reference for the purpose of plotting amounts of hybridization of each solution nucleic acid, reference and test sample, to each of the immobilized cloned nucleic acids. The results are plotted as a function of the linear

20   position of each of the cloned immobilized nucleic acids on a chromosome. Then a representation is made of a portion or of an entire chromosome, or of a plurality of chromosomes, or of a complete set of chromosomes (autosomes with or without sex chromosomes), i.e., of the entire genome. Results obtained from analyzing both sets of data are combined to reveal changes that would otherwise be undetectable if label reversal was not

25   used. This is because small fluctuations from a ratio of 1.0 become statistically significant when the dye swap data are plotted together, which might not be significant if only a single mixture was used.

Further, it is desirable to compare multiple test subjects with the same reference sample. In any of these uses, multiple iterations of identical arrays on a single surface are

30   highly advantageous or necessary.

Prior to the methods and surfaces as described herein, it has been necessary to conduct such analyses using a plurality or a multiplicity of replicas of the array of elements printed on each of several different surfaces, each replica having one complete iteration of all of the elements within the array, or possibly having two iterations with elements interspersed,

or possibly having two iterations closely aligned to eachother, the two iterations serving as statistical replicas for improved accuracy of a single hybridization mixture. It has not been possible to contact a single surface successfully with two different mixtures and maintaining a separate integrity of each of the different mixtures. For example prior to the present

5   methods, a dye swap analysis was performed with two mixtures, the first being a mixture of the test nucleic acid labeled with the first dye and mixed with the reference nucleic acid labeled with the second dye, and the second being the test nucleic acid labeled with the second dye mixed with the reference nucleic acid labeled with the first dye, the two mixtures were then analyzed using separate arrays on each of two different surfaces.

10   The use of a multiplicity of different surfaces for separate hybridizations of each of different nucleic acid mixtures can be a source of variability, e.g., in efficiency of binding of spots to each surface, hybridization due to variability in conditions, minor variations in concentration of each nucleic acid, variation in concentrations, different efficiencies in elution of non-specifically bound materials due to minor variations in washing procedures or

15   solutions, at the time of hybridization to each separate surface, or variations in photomultiplier settings in a scanner used to visualize and evaluate the array, after hybridization to each separate surface. Accordingly, the present surfaces provided herein address this problem in the prior common usage by having multi-arrays, which are a plurality of the arrays present together on a single surface.

20   In a non-limiting example, two arrays are located at distal ends of a planar substrate such as a standard glass microscope slide, however alternative shapes and sizes of substrates, and shapes and sizes of arrays, are within the scope of surfaces, kits and methods envisioned herein. For example, a substrate may be a one inch by 3 inch microscope slide, and may have a plurality of arrays such as two arrays, one at either end, or four arrays in a linear

25   arrangement. A larger substrate such as a square slide may have four arrays, one in each corner, or nine arrays with three arrays on each side and one in the middle.

Further, barriers to maintain separation of fluids deposited on each array during hybridization may be used, the barriers being placed between each of the arrays, in addition to embodiments of the surface in the absence of barriers, as described herein. The barriers

30   are physical "dykes" or "dams" having a height above the plane of the substrate face or surface, and such barriers include raised portions of the substrate as manufactured, or as added subsequently. Alternatively, the barriers may be hydrophobic materials that are printed on the substrate to produce a "strip" which can prevent the flow of an aqueous solution from

one array to another. The barriers can be added before or after printing or depositing the
micro-arrays, to produce the multi-array surfaces.

The barriers are comprised of a material that is not soluble in aqueous solution, and
the material hydrophobic. Exemplary hydrophobic materials for barrier construction or
5   printing include: polyethylene, silicone, paraffin, and Teflon®.

Hybridization using the "multi-array surfaces" having multiple arrays on a single
surface of a single substrate, is conducted by adding each of the hybridization mixtures to an
iteration of the array and protecting the hybridization mixture with a cover to prevent loss of
volume of solution by evaporation. The cover further acts to confine each hybridization of a
10   particular sample or mixture of samples, labeled with one or more dyes as described above, to
the appropriate micro-array. A pre-determined amount of hybridization mixture is deposited
above each of the arrays, such that addition of a cover, for example placed directly on the
fluid, yields a resulting thin layer of fluid above the array in which the sample nucleic acids
can hybridize to complementary sequences within the array. Hybridization for each array on
15   the surface is conducted under a separate cover.

Conditions for hybridization can be modified, for example, the hybridization solution
can be altered, to improve and assure fluid separation of the multiple hybridizations on the
surface. For example, viscosity of the hybridization fluid may be increased to reduce fluidity
by adding one or more solutes that do not interact with the nucleic acids during the
20   hybridization. Exemplary solutes include small molecules that are viscous liquids such as
glycerol; polymers of small molecules such as sugars, exemplified by but not limited to
dextrans; and starches such as corn starch; polymers of amino acids which are synthetic
polypeptides or naturally occurring proteins such as albumins and gelatins; and synthetic
polymers, for example, polyethylene glycol, or polyacrylamide or agarose, each solute at a
25   concentration sufficient to increase viscosity without significantly affecting mobility of the
dissolved nucleic acids for interaction and hybridization (annealing to form a double stranded
complex) to the immobilized nucleic acids. The viscosity increasing solute may be
chemically modified to improve its properties, for example, to render it resistant to digestion
by extracellular enzymes of bacteria and fungi. Solutions for hybridization may be stored
30   with antibiotic or growth inhibiting materials to retard spoilage during storage; alternatively,
solutions may be frozen or lyophilized for convenient storage for later use.

The multi-array surfaces and methods herein are not limited to performance of dye
swap analyses. For example, a multi-array surface having two or more, e.g., four, five, six or
even nine iterations of an array can be used to analyze multiple samples, for example, a

plurality of members of a nuclear family, or multiple siblings and a proband carrying a chromosomal disorder, which can now be analyzed together on a single substrate having multiple micro-arrays, using separate hybridizations. Further, any multiple number of subjects can be analyzed simultaneously on a single substrate, or any one subject can be

5   analyzed in mixtures of different reference samples. Different reference samples can be prepared in advance from each of relevant different species for extensive repeated use as a standard of comparison with multiple different test samples, or from specific animal strains having one or more of several different known transgenes or mutations, or different predetermined single BAC clone nucleic acid or mixtures of nucleic acids from two or more

10   BAC clones.

Chromosomal analysis using calibration spots and disease-negative clones

Calibration spots that act as positive controls for hybridization of a sample, and that are located within an array have been described (see, U.S. patent application 2003-0186250-A1, published Oct. 2, 2003, and incorporated herein in its entirety by reference). In

15   embodiments of the surfaces and methods provided herein, calibration spots may include a subset of cloned nucleic acids, for example, those clones of the human genome carrying sequences not known by any published references to be associated with a chromosomal disorder or disease. The term, "non-reactive" as applied to a specific cloned sequence of nucleic acid of known chromosomal location, means that the nucleic acid generally

20   hybridizes to a full extent to a genomic nucleic acid from any test subject, i.e., and is "non-reactive" because it does not give a false "positive" diagnosis of a chromosomal disorder. Because the non-reactive or backbone clones are positive controls for hybridization, they are therefore expected to be non-reactive with a test sample for detection of a chromosomal disorder.

25   A calibration spot may be a mixture of nucleic acids from any combination of other elements present in the array, or can be a mixture of a subset of such elements, or can be a nucleic acid not so represented in the array. For example, a calibration spot can be a mixture of backbone clones as defined above, for any one syntenic set of clones representing the syntenic chromosome as chosen by the user, or for all of the chromosomes in the human

30   genome or in a genome of any other organism. An exemplary calibration spot may comprise a mixture of nucleic acids, for example, from backbone clones, for example, from about 10, from about 20, from about 40, or from about 80 clones such as backbone or non-reactive clones. An exemplary but non-limiting calibration spot contains 72 non-reactive backbone clones, selected to represent nucleic acid from each of the set of human autosomes and sex

chromosomes. An alternative calibration spot contains nucleic acid from an unrelated heterologous species, such as a fish or amphibian, for purposes of standardizing hybridization, in which case an internal control carrying a recognizable label can be added or "spiked" into each hybridization mixture of a test sample and a reference sample.

5   Representation of each chromosome is made by calculating ratios of labels in each of the two double dye-labeled hybridizations (dye swap) and relative amounts are plotted graphically as a function of distance of each cloned chromosomal portion from the p terminus conventionally shown on the left. By convention, one of the two double labeled materials is plotted in a consistent color (e.g., red), and the other in a different color (e.g., blue), such that

10   deletion of a portion of nucleic acid in a test subject is displayed in red above the 1.0 ratio line (see Figs. 1 and 2), and an insertion such as an amplification is plotted as blue above the 1.0 ratio line.

In addition, the arrays provided herein as shown in drawings and examples herein, include cloned nucleic acids from portions of each chromosome that are not associated with

15   any known chromosomal disorders, so that representations of a chromosome of a test subject's DNA is facilitated, and a chromosomal disorder on a given chromosome is more readily distinguished from normal portions of that chromosome.

EXAMPLES

20

The following examples are offered to illustrate embodiments of the invention, and are not to be construed as further limiting. The methods are used throughout the examples.

Example 1. Making BAC Clone Nucleic Acid Arrays

25   BAC clones (Tart of et al., 1987, CA Cetheda Res. Lab Focus 86:184; available from the Bio Laboratories, Carlsbud CA) containing inserts of greater than thirty kilobases (30 kb), and up to about 300 kb, are grown up in Terrific Broth medium (commercially available from numerous suppliers). Large inserts, e.g., clones >300 kb, and small inserts, about 1 to 20 kb, can also be used. DNA is prepared by a modified alkaline lysis protocol (see, e.g.,

30   Sambrook). A genomic DNA sample used in labeling experiments is prepared by protocol to be substantially free of RNA and proteins. Any general protocol for this purpose can be used, and the following procedure is not to be construed as limiting.

Following lysis of cells and removal of cell debris, ribonuclease (DNase-free, 10 mg/ml) is added to the DNA sample to a final concentration of 20-100 microg/ml, and the

mixture is incubated at 37°C for 30 min. Proteinase K is added at a final concentration of 100 microg/ml, and the sample is incubated at 50°C for one hour. The sample is cooled to room temperature, and an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) is added. The two phases are gently mixed by rotation on a wheel or end-to-end turning, for 10 min,

5   and the phases are separated by centrifugation at 10,000g for 3 min.

The aqueous layer is removed, and is similarly re-extracted until no interface material is observed. Chloroform is used to remove remaining phenol, the mixture is again centrifuged, the aqueous layer is removed to a clean tube, and DNA is precipated by adding a one-twelfth volume of 5M NaCl. The solutions are mixed by slow end-to-end turning,

10   followed by addition of 2.5 volumes of ice cold 100% ethanol (or a 0.75 volume of room temperature isopropanol). After addition of ice-cold ethanol, the sample is incubated at -20°C for 30 min to one hour (or about 15-30 min at room temperature if isopropanol is used).

The DNA precipitate is collected by micro-centrifiguration at maximum speed for 10 min. The ethanol (or isopropanol) supernatant is carefully removed, and 1ml of 70% ethanol

15   is added to the pellet to remove precipated salt. The 70% ethanol is gently removed, and a second rinsing of the pellet with 70% ethanol is performed. The second 70% ethanol is removed, and the pellet is dissolved in sterile distilled water. A DNA concentration of 100-200 nanograms/microliter is obtained, the DNA having an average molecular weight greater than the 8,454 base pair lambda DNA-BstE II digested marker, and substantially free of

20   RNA. The DNA is labeled as described herein.

The DNA is then chemically modified as described by U.S. patent number 6,048,695. The modified DNA is then dissolved in proper buffer and printed directly on clean glass surfaces as described by U.S. patent number. 6,048,695. Usually multiple spots are printed for each clone. Two or more iterations or sets of each complete array are printed on the

25   surface, each complete array separated by a barrier, or separated by a space having no spots, i.e., a plurality of non-contiguous arrays.

Example 2. Labeling of Genomic DNA

Genomic DNA for test and reference samples substantially is prepared substantially

30   as above, and is substantially free of RNA and protein. The DNA is pretreated to obtain small more uniform pieces, is differentially labeled, and is hybridized to a slide having an array spotted as described herein, washed, and scanned and analyzed for detection of chromosomal abnormalities.

Pretreatment includes digestion with a DNase, preferably a four base pair cutter such as *Eco*RI, to reduce the size of the genomic DNA. About one microgram of sample DNA is incubated with *Eco*RI (2 microliters or 20 units) for about 16h at 37° C. Extent of completion of the reaction is analyzed by electrophoresis in 0.8% Agarose, and is determined to be

5   complete if a relatively homogeneous smear from 600 bp to ≥ 20 kb is observed. If the digestion is complete, the reaction is terminated by incubating the tube at 72° C for 10 min. DNA is re-purified by phenol/chloroform extraction and ethanol precipitation as described herein, or by equivalent means (Zymo Research kit, DNA Clean and Concentrator ™-5, Cat. No. D4005, Hornby, Ontario, Canada).

10   An aliquot of each of a test sample and a reference sample DNA is labeled with each of Cy3™-dCTP and Cy5™-dCTP, to facilitate co-hybridization. About 500 micrograms of re-purified digested DNA in 25 microliters of water is added to 20 microliters of 2.5X random primer reaction mix (Gibco/BRL BioPrime labeling kit, Gibco/BRL, Bethesda MD). Samples are mixed, and incubated at 100°C for 5 min, then incubated on ice for 5 min. To

15   each tube of DNA, 2.5 microliters of Spectral Labeling Buffer (Spectral Genomics, Inc., Houston, TX), and 1.5 microliters of Cy3™-dCTP (1 mm stock) or Cy5™-dCTP (1 mm stock) is added. Then 1 microliter of Klenow fragment (from the Gibco/BRL BioPrime Labeling kit) is added, the solution is mixed and re-collected by brief centrifugation, and the sample is incubated at 37°C for about 1.5 to 2 hours. Samples are placed on ice, and

20   analyzed by electrophoresis (using 0.8% Agarose) to determne probe size, which should have a range of about 100 bp to about 500 bp. The reaction is terminated by addition of 5 microliters of 0.5M EDTA, pH 8.0, and incubated at 72°C for 10 min, followed by placing the samples on ice. Tube contents are ready for hybridization, or can be stored at -20°C until required.

25   For hybridization, two mixtures of labeled test sample and labeled reference sample are prepared. Tube contents of Cy3™-labeled test sample DNA is mixed with tube contents of Cy5™-labeled reference DNA, and tube contents of Cy3™-labeled reference DNA is mised with tube contents of Cy5™-labeled test DNA. Spectral Hybridization Buffer (45 microliters; Spectral Genomics, Inc., Houston, TX) is added to each of the tubes, and the

30   contents are precipitated by adding 11.3 microliters of 5M NaCl and 110 microliters of room temperature isoproponol. Samples are mixed and incubated in the dark at room temperature, for about 10-15 min, centrifuged at maximum speed for 10 min preferably in the dark, and supernatants are aspirated. Pellets are rinsed with 500 microliters of 70% ethanol and air-dried briefly in the dark, 10 microliters of sterile water is added, and tubes are incubated at

room temperature for 5 min and thoroughly mixed. When pellets are dissolved, 30 microliters of Spectral Hybridization Buffer II (Spectral Genomics Inc., Houston, TX) are added, and tube contents are mixed by repeated pipetting. Samples are denatured by incubation at 72° for 10 min.

5

Example 3. <u>Optimization of Printing of Nucleic Acid Spots as a Function of Ionic Strength</u>

Spots are generally printed robotically on a glass slide in a pattern, in duplicate blocks, each block containing hundreds of spots. Following drying, each spot can be examined microscopically to assess quality of deposition of the biological materials. High

10 quality is indicated by a smooth, uniform appearance of a high proportion of the printed spots, and is reflected in a uniform pattern of hybridization. Low quality is indicated by spots that appear as "mountains" such that sample is deposited non-uniformly in the center, or as several "hills" or in an X-shaped configuration across diameters of the spot, and is reflected in a similarly distributed hybridization.

15

To determine the effect of ionic strength of printing buffer on the quality of the spots in arrays on glass slides, six samples of BAC DNA were prepared as shown in Table 1.

Table 1. Buffer component concentrations of experimental printing buffers, in millimolar*

Component:	Tube #: 1	2	3	4	5	6
NaOH	25	51.25	77.5	104	130	156.25
TRIS-HCl	50	102.5	155	208	260	312.5
EDTA	5	10.3	15.5	20.8	26	31.2

20 * All buffers were prepared from standard stock solutions of 400 mm NaOH and 1M TRIS-HCl, 0.1M EDTA, pH 4.3.

An aliquot of DNA was added to each of tubes 1 through 6, and spots were printed on glass slides and were dried by standard procedures.

25 Examination of printed slides showed best quality, i.e., uniform distribution of DNA deposited on the greatest proportion of printed spots, of those printed at conditions found in tube #3; slides with spots printed in buffers with tubes #1, #2, #4 and #5 were of poorer quality appearance and gave poorer quality of hybridization than did those printed with buffer of tube #3. These intermediate ionic strengths in tubes #2-5 have a higher than ionic strengths

30 than the previously used buffer (tube #1). Optimum appearance of the printed spot was reflected in greatest reliability and reproducibility of hybridization of reference samples, and was seen with spots printed in the buffer of tube #3.

Production of buffer for printing of spots was, as a result of data shown above, changed for further printing by mixing equal volumes of two stock solutions: solution I is 150 mM NaOH, and solution II is 300 mM TRIS-HCl, 30 mM EDTA, pH of 4.3. Final concentrations of components of this printing buffer are: 75 mM NaOH, 150 mM TRIS-HCl,

5   and 15 mM EDTA.

Example 4. Preparation of Syntenic Arrays

DNA is prepared as above from characterized BAC clones having DNA inserts that are syntenic for each mouse chromosome and for each human chromosome of interest. The

10   DNA is deposited (printed) in an array of addressible locations (spots) on a glass slide. The arrayed DNA can be from a normal chromosome of a human and a normal chromosome of any other species such as a mouse, or from a chromosome of a human having a known disease, or a mouse having a mouse disease that is a model of a human disease, such as lung cancer, mesothelioma, adenocarcinoma, and prostate cancer. The presence both of human and

15   another species, i.e., mouse syntenic sequences that are homologous to the human sequences, constitutes a syntenic array.

Syntenic arrays are printed also for other combinations of species DNA, such as human-rat; human-chimpanzee; human-dog, and the like.

Test and reference samples (positive and negative controls) are prepared from

20   genomic DNA of diseased and normal subjects, not necessarily limited to the subjects of those species identified by having the array of immobilized DNA. Each is labeled with a first and a second fluorescent dye, and two mixtures of test and reference samples are made. Each of the two mixtures is separately hybridized to an array of the multiple arrays on the surface.

25   Example 5. Detection of Chromosomal Abnormalities using Ratio Plot Analysis

Fig. 1a is a ratio plot of a sample of patient DNA hybridized to immobilized BAC clones of chromosome 18, using a linear display of ratio of DNA from the 25 BAC clones, in order from left to right of p to q arms. Two iterations of the hybridization were performed, one in which the hybridization was performed with a mixture of test DNA (from the patient)

30   labeled with Cy3™ and reference normal DNA is labeled with Cy5™; and a second iteration in which the with a mixture of test DNA (from the patient) labeled with Cy5™ and reference normal DNA is labeled with Cy3™. After data are analyzed, the ratio of Cy5:Cy3 is determined and is plotted as shown in the Figs.

At the q terminus of chromosome 18, the patient data indicates a deletion, as both normalized functions are increasing in value for the same three q-terminus BAC hybridizations (both show a simultaneous deviation from a modal value of 1.0, regardless of whether the test sample or reference sample is the numerator of the ratio). A similar ratio plot

5   shown in Fig. 1b indicates that DNA at the q-terminus of chromosome 4 the same patient carries an insertion, as both functions deviate in opposite directions for two BAC clones, with the blue (reference sample in numerator) DNA being greater than the red (test sample DNA in numerator) hybridization.

Fig. 2 shows a ratio plot of the X chromosome of a different male patient, which by

10   the computations described herein indicates that there is an insertion (amplification) of sequences found at the p-ter of this chromosome. The amplification is shown to extend over at least three BAC clones. The deviations shown at the q-ter of this chromosome are considered random and not significant, as only one of the functions deviates to any extent from a modal value of 1.0.

15

Example 6. Testing of Cancer Tissue for Chromosomal Abnormalities Using Syntenic Arrays

Cancer cells of strains of mice having particular cancer conditions, such as lung cancer, mesothelioma, adenocarcinoma, and prostate cancer, are used to make DNA test samples. Test samples are prepared from tissue obtained from individual mice having

20   different stages of each of the diseases. The DNA is fragmented, labeled, and hybridized as described herein. Abnormalities associated with each of these conditions are determined from ratio plots of hybridization data, using a syntenic array having probe elements for each of human and mouse syntenic sequences from cloned BAC chromosomes that are highly homologous in humans and mice.

25   Presence, extent, and location of these abnormalities are then determined using test sample DNA obtained from human patients having various stages of the disease, such that a diagnostic instrument capable of diagnosing and prognosing the stage of the cancer is established. The data show not only the location of the cancer-induced changes on the mouse chromosome set, but also show the location of such changes that would occur in humans, and

30   at loci for homologous sequences on the human chromosome set.

Example 7. Toxicological Testing With Syntenic Arrays

Groups of test animals (or test cells), for example, rodents such as rates, mice, or guinea pigs, are exposed to a chemical. In one experiment, each test animal member of a

group is injected with a concentration of one or more chemicals to be tested for genotoxic or

mutational activities, with different groups receiving each of different doses, and one group

receiving only carrier or solvent (negative control), so that a range of concentrations of the

composition is tested to establish a dose response curve. A genotoxic activity is detected as

5   causing one or more of deletion/insertion mutations, or translocation of DNA from one locus

and chromosome to another locus on the same or a different chromosome, or amplification of

a region of a chromosome. In another experiment, cells in culture are exposed to a range of

each of the chemicals, individually, or in combinations such as in groups of 10 chemicals in a

single tube, i.e., related chemicals are tested in sibling groups which can be further analyzed

10   individually or in smaller groups to obtain a correlation with genotoxicity.

        Following exposing or contacting the cells or organisms with the one or more

chemical compositions for varying extents of time and/or concentrations, DNA is prepared

from a cell sample, e.g., from a somatic tissue from an autopsy or biopsy, e.g., from ovarian

or testicular tissue such as a blood sample, or from the cells in culture that were exposed to

15   the chemical. The DNA is fragmented and labeled as described above with each of two

fluorescent dyes or equivalent detectible labels, for test sample to be mixed with oppositely

labeled reference sample, and the mixtures are each hybridized to an iteration of a syntenic

array having both the rodent, e.g., mouse and human genomes, or cloned nucleic acids from

particular chromosomes, arrayed as probe elements. The data show extent and location of

20 · DNA damage in the test sample from test animals or cells, both in the rodent chromosome

set, and for homologous sequences on the human chromosome set.

        A number of embodiments of the invention have been described. Nevertheless, it will

be understood that various modifications may be made without departing from the spirit and

scope of the invention. Accordingly, other embodiments are within the scope of the following

25   claims.

Example 8. Syntenic Arrays for Analyzing Chromosomal Abnormalities in a Transgenic
Animal

        A large variety of transgenic animals are available for research and testing purposes.

30   An animal lacking a gene function, in this example, a mouse strain having a disruption in a

gene encoding nitric oxide synthase (NOS), can be used for screening purposes, to identify

compositions capable remediating a phenotype (U.S. patent number 6,310,270 issued Oct. 30,

2001).

A DNA sample is prepared from a blood sample of each animal in treated and control groups, and is labeled with each of two fluorescent dyes as described in examples above, and mixtures of labeled samples and differently labeled reference DNA are prepared as above. Treated and control animal DNA mixtures are hybridized to the syntenic chips having each of

5   human and mouse BAC cloned DNA in each array block.

Results indicate which of those compositions that are active in remediating the disrupted phenotype, NOS, that have not caused chromosomal abnormalities in the test animals. It is envisioned that the majority of compositions do not cause chromosomal abnormalities, so an additional positive control group of animals administered an agent

10   known to cause chromosomal abnormalities, such as benzene, is included. In this example, compositions are screened both for a pharmacological activity, remediation of lack of NOS, and for induction of chromosomal abnormality, i.e., teratogenicity and mutagenicity, in the same groups of test and control animals. The syntenic chip readout contains mouse and human elements, and as it is envisioned that any chromosomal abnormality observed in the

15   mouse in comparison to mouse reference DNA has occurred de novo, the chromosomal abnormality is simultaneously analyzed both with respect to the mouse chromosome and genome, and to homologous elements of the human genome on the single surface of the syntenic chip.

20   Example 9. Syntenic Arrays for Analyzing Chromosomal Abnormalities in an Animal Model of Human Disease

Groups of non-obese diabetic (NOD) mice are administered a each of variety of compositions, to determine whether these are capable of remediating or preventing type I or insulin-deficient diabetes. Agents are administered intravenously.

25   At the end of the treatment protocol, a DNA sample is prepared from a blood sample of each animal in treated and control groups, and is labeled with each of two fluorescent dyes as described in examples above, and mixtures of labeled samples and differently labeled reference DNA are prepared as above. Treated and control animal DNA mixtures are hybridized to syntenic chips having each of human and mouse BAC cloned DNA in each

30   array block.

Results indicate which of those compositions that are active in remediating the model disease, diabetes, that have not caused chromosomal abnormalities in the test animals. It is envisioned that the majority of compositions will not cause chromosomal abnormalities, so a positive control group of animals administered an agent known to cause chromosome

45

abnormalities can be included in the example. In this manner compositions have been screened both for a pharmacological activity, remediation of diabetes, and for mutagenicity, in the same groups of test and control animals. The syntenic chip readout contains mouse and human elements, and as it is envisioned that any chromosomal abnormality observed in

5   the mouse in comparison to mouse reference DNA has occurred de novo, the chromosomal abnormality is simultaneously analyzed both with respect to the mouse chromosome and genome, and to homologous elements of the human genome on the single surface of the syntenic chip.